ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

ESTIMACIÓN DEL RENDIMIENTO POR HECTÁREA DE MAÍZ DURO SECO EN EL ECUADOR MEDIANTE REDES NEURONALES ARTIFICIALES

TRABAJO PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO MATEMÁTICO

PROYECTO DE INVESTIGACIÓN

JENDRY PAÚL TOAPANTA GUAMÁN
jendry_toapanta@outlook.com

DIRECTORA: PHD. ADRIANA UQUILLAS ANDRADE
adriana.uquillas@epn.edu.ec

QUITO, MAYO 2019

DECLARACIÓN

Yo JENDRY PAÚL TOAPANTA GUAMÁN, declaro bajo juramento que el trabajo aquí escrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedo mis derechos de propiedad intelectual, correspondiente a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su reglamento y por la normatividad institucional vigente.

Jendry Paúl Toapanta Guamán

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por JENDRY PAÚL TOAPANTA GUAMÁN, bajo mi supervisión.

Directora del Proyecto
PhD. Adriana Uquillas Andrade

AGRADECIMIENTOS

A mi madre, a mi padre, a mis hermanas y hermanos, gracias por el apoyo constante y acompañarme a lo largo de mi carrera.

A la Dra. Adriana Uquillas, por su paciencia, consejos y sugerencias brindadas para el desarrollo y culminación de este trabajo.

DEDICATORIA

A mi madre María Rosa, por su amor abnegado y apoyo incondicional. A mi padre Jorge, por confiar en mi y no dejar que me rinda.

Índice general

Índice de figuras III

Índice de cuadros V

Resumen VII

Abstract VIII

1. Introducción **1**

1.1. El maíz duro seco en el contexto mundial 1

1.2. El maíz duro seco en el Ecuador 2

1.3. Operativo de maíz duro seco 5

2. Marco teórico **8**

2.1. Regresión lineal múltiple 8

2.1.1. Modelo de Regresión Múltiple 8

2.1.2. Estimación 10

2.1.3. Validación 15

2.1.4. Medidas de bondad de ajuste 26

2.2. Redes neuronales artificiales 28

2.2.1. Neurona biológica 28

2.2.2. Neurona artificial 29

2.2.3. Arquitecturas de Redes Neuronales Artificiales y Procesos de Aprendizaje 33

2.2.4. Redes prealimentadas de una capa 39

2.2.5. Redes prealimentadas multicapa 46

2.2.6. Versiones optimizadas del algoritmo backprop 55

3. Análisis de los datos y clusterización **58**

3.1. Descripción de la base de datos 58
3.1.1. Selección de las variables 58
3.1.2. Clusterización por provincias, cantones y variedad 63

4. Desarrollo de un Modelo de Regresión Lineal Múltiple y un Modelo de Redes Neuronales Artificiales 70

4.1. Desarrollo de un modelo de regresión lineal múltiple 70
4.1.1. Modelo de RLM para el ciclo invierno 72
4.1.2. Modelo de RLM para el ciclo verano 77
4.2. Desarrollo de un modelo de redes neuronales 82
4.2.1. Red neuronal para el ciclo de invierno 83
4.2.2. Red neuronal para el ciclo de verano 89

5. Conclusiones y recomendaciones 96

6. Anexos 100

Bibliografía 102

Índice de figuras

1.1. Producción y rendimiento de maíz duro seco 3
1.2. Consumo aparente nacional, producción nacional e importación 4
1.3. Importaciones de maíz duro seco 4
1.4. Exportaciones de maíz duro seco 5

2.1. Estructura de una neurona biológica 28
2.2. Estructura de una neurona artificial 29
2.3. Funciones de activación parcialmente diferenciables 32
2.4. Funciones de activación totalmente diferenciables 34
2.5. Red prealimentada de una capa 36
2.6. Ejemplo de una red prealimentada multicapa 37
2.7. La red Perceptrón 39
2.8. Red ADALINE 42
2.9. La función de error cuadrático medio con respecto al número de iteraciones 45
2.10. Ilustración de las dos etapas de una red MLP 48
2.11. Notación usada en el algoritmo backpropagation 48
2.12. Ilustración de una neurona usada en el algoritmo backpropagation 49
2.13. Representación de las conexiones neuronales para ajustar los pesos de las neuronas de la segunda capa 52
2.14. Representación de las conexiones neuronales para ajustar los pesos de las neuronas de la primera capa 54

3.1. Rendimiento promedio por provincia y ciclo 64
3.2. Suma del error interclúster provincial para los distintos valores de k elegidos 65
3.3. Rendimiento promedio por cantón y por ciclo 66
3.4. Suma del error interclúster cantonal para los distintos valores de k elegidos 66
3.5. Rendimiento promedio por variedad y por ciclo 68

3.6. Suma del error interclúster de variedad para los distintos valores de k elegidos 68

4.1. Matriz de correlación del modelo de invierno 74
4.2. Gráfico qq de probabilidad normal de los residuos del modelo de invierno . 75
4.3. Gráfico de los residuos del modelo de invierno 76
4.4. Matriz de correlación del modelo de verano 79
4.5. Gráfico qq de probabilidad normal de los residuos del modelo de verano . . 80
4.6. Gráfico de los residuos del modelo de verano 81
4.7. Red perceptrón multicapa para el ciclo de invierno 84
4.8. Modelo de red neuronal para el ciclo de invierno 90
4.9. Red perceptrón multicapa para el ciclo de verano 91
4.10. Red neuronal ajustada para el ciclo de verano 95

Índice de cuadros

1.1. Principales datos de maíz duro seco a nivel mundial 2

2.1. Niveles y sus valores críticos para la prueba de Anderson-Darling 26

3.1. Temperatura promedio ciclo 59
3.2. Precipitación acumulada ciclo 60
3.3. Fertilización aplicada 60
3.4. Agricultores que recibieron el Kit de beneficio 61
3.5. Capacitación 61
3.6. Agricultores miembros de asociaciones 62
3.7. Notación de las variables obtenidas de la base de datos 63
3.8. Clusterización por provincias 65
3.9. Clúster por Cantones 67
3.10. Clúster por variedad 69

4.1. Modelo de RLM de invierno 73
4.2. Prueba RESET del modelo de RLM de invierno 73
4.3. Pruebas de normalidad del modelo de invierno 75
4.4. Test de homocedasticidad del modelo de verano 76
4.5. Modelo de RLM de verano 78
4.6. Prueba RESET del modelo de verano 79
4.7. Pruebas de normalidad del modelo de verano 80
4.8. Test de homocedasticidad del modelo de verano 81
4.9. Arquitecturas para el ciclo invierno con el Algoritmo Backpropagation . . . 85
4.10. Arquitecturas para el ciclo invierno con el Algoritmo Rprop+ 86
4.11. Arquitecturas para el ciclo invierno con el Algoritmo Rprop- 87
4.12. Mejores arquitecturas para el ciclo de invierno 87

4.13. Mejores arquitecturas para los datos de todo el ciclo de invierno 88
4.14. Comparación de los modelos ANN y RLM para el ciclo invierno 88
4.15. Matriz de pesos del modelo de invierno 88
4.16. Arquitecturas para el ciclo verano con el Algoritmo Backpropagation . . . 91
4.17. Arquitecturas para el ciclo verano con el Algoritmo Rprop+ 92
4.18. Arquitecturas para el ciclo verano con el Algoritmo Rprop- 93
4.19. Mejores arquitecturas para el ciclo de verano 93
4.20. Comparación de los modelos ANN y RLM para el ciclo verano 94
4.21. Matriz de pesos del modelo de verano 94

Resumen

El maíz es una gramínea considerada de alto poder económico y de seguridad alimenticia debido a su uso en la alimentación humana y en la elaboración de balanceados en la industria pecuaria en el Ecuador. En el presente trabajo el rendimiento por hectárea de maíz es modelado mediante regresión lineal múltiple y redes neuronales artificiales. El rendimiento fue modelado en función de los parámetros de fertilización, rendimiento, ubicación geográfica y variables climáticas. tiene como objetivo determinar un modelo de redes neuronales artificiales para estimar el rendimiento de maíz duro seco por hectárea para cada ciclo de producción, invierno y verano, mediante el uso de variables climáticas, de fertilización, de ubicación geográfica del cultivo y variedad de la semilla. Para esto, se determinan los factores influyentes en el rendimiento del cultivo para cada ciclo y se plantea un modelo de regresión lineal múltiple y un modelo de redes neuronales. Finalmente, se compara los resultados obtenidos en ambos modelos mediante estadísticos de bondad de ajuste.

Palabras clave: Redes neuronales artificiales, rendimiento de maíz en el Ecuador, algoritmo Backpropagation, algoritmo Resilient-Propagation, regresión lineal múltiple.

Abstract

Maize is an important grass in the economy and food security due to its use in human food and as a feed in the livestock industry in Ecuador, which motivates the study of the factors that influence the yield of this crop. The objective of this work is to determine a model of artificial neural networks to estimate the yield of maize per hectare for each production cycle, winter and summer, through the use of climatic variables, fertilization, geographic location of the crop and variety of the seed. For this, the influential factors in crop yield are determined for each cycle and a multiple linear regression model and a neural network model are proposed. Finally, the results obtained in both models are compared using goodness-of-fit statistics.

Keywords: Artificial neural networks, maize yield in Ecuador, Backpropagation algorithm, Resilient-Propagation algorithm, multiple linear regression.

Capítulo 1

Introducción

En este capítulo se presenta un panorama global de la producción de maíz duro seco. En la primera sección se describe las estadísticas a nivel mundial de los principales productores y consumidores de maíz. Luego se presenta la evolución de este cultivo a nivel nacional. Finalmente se describe el Operativo de Rendimientos Objetivos, programa a partir del cual se obtuvieron los datos para este trabajo.

1.1. El maíz duro seco en el contexto mundial

El maíz (*Zea mays*) es uno de los granos alimenticios más antiguos que se conocen, pertenece a la familia de las Poáceas. Requiere de precipitación de 550mm a 2000mm por año, una temperatura entre 24°C a 28°C, una altitud de 45 a 125msnm y radiación solar de 750 a 100 horas luz por año. Entre los principales usos se encuentran la alimentación humana, animal, combustible de etanol, edulcorantes, jarabe de maíz, producción de almidones, entre otros [31].

En el Cuadro 1.1 se presentan los diez países con mayor producción a nivel mundial y que concentran el 85 % del total, destacando países como Estados Unidos y China con una producción equivalente al 36 % y 21 % respectivamente para el período $2017-2018$, según datos del Departamento de Agricultura de los Estados Unidos (USDA) [31]. Comparando la producción con los niveles de consumo se tiene que los países de México, China y la Unión Europea no son autosuficientes, es decir, su producción de maíz duro seco no abastece su consumo local.

Respecto al comercio de esta gramínea, el mayor exportador es Estados Unidos con

el 32 % de las exportaciones, seguido de Brazil y Argentina con el 23 % y 18 % respectivamente. En lo referente a las importaciones México, la Unión Europea y Japón son los principales importadores de maíz duro seco; en Sudamérica se tiene a Colombia como el mayor importador con el 3 % de las importaciones mundiales.

Cuadro 1.1: Principales datos de maíz duro seco a nivel mundial

País	Producción (millones Tm)	País	Consumo (millones Tm)	País	Exportación (miles Tm)	País	Importación (miles Tm)
Estados Unidos	371	Estados Unidos	318,7	Estados Unidos	48,5	México	16,5
China	215,9	China	240	Brazil	35.5	Unión Europea	16
Brazil	95	Unión Europea	74,8	Argentina	27	Japón	15
Unión Europea	60,1	Brazil	61,5	Ucrania	20,5	Egipto	10
Argentina	42	México	42,3	Rusia	4,8	Korea del Sur	9,7
México	26,2	India	25,8	Paraguay	2.5	Vietnam	9,5
Ucrania	25	Egipto	15,9	Unión Europea	2	Irán	9
India	25	Japón	15,1	Sudábrica	1,9	Colombia	5
Canada	14,1	Canada	14	Birmania	1.2	Argelia	4,7
Rusia	13.5	Vietnam	13,4	México	1.3	Taiwan	4,5
Otros	156.8	Otros	245,2	Otros	6,2	Otros	50,9
Total	1,044.6	Total	1.066,7	Total	151,4	Total	150,8

Fuente: USDA, FAS Grain: World Markets and Trade
Elaboración: Autor

1.2. El maíz duro seco en el Ecuador

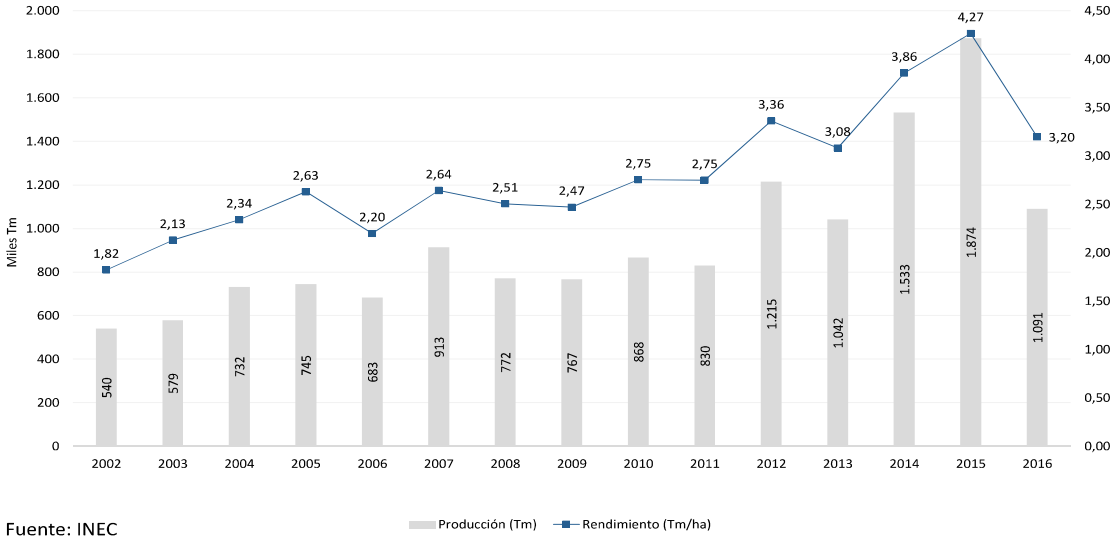
En el Ecuador el maíz duro seco es uno de los cultivos más importantes de la economía nacional. Constituye la principal materia prima para la elaboración de balanceados destinados a la industria animal, especialmente a la avicultura comercial. Se lo cultiva en 21 provincias y de forma mayoritaria y extensiva en Los Ríos, Manabí, Guayas, El Oro y Loja. De acuerdo al III Censo Agropecuario del 2000 en el país existían aproximadamente 82 mil unidades productoras destinadas al cultivo y producción de maíz duro seco.

En el período comprendido entre el 2002 al 2016 la variación de la superficie sembrada ha sido notoria, con un crecimiento anual promedio del $1,8\,\%$, y un crecimiento del $15,2\,\%$ en el 2016 respecto del 2002. Las cifras de producción y rendimiento[1] para le período 2002 al 2016 se presentan en la figura 1.1, se observa que los valores de producción se han duplicado de 540 mil toneladas en 2002 a 1.091 mil toneladas en 2016; cabe destacar que en 2015 fue el año de mejor producción, llegando a 1.874 mil toneladas. Los rendimientos promedio nacionales han pasado de 1.82 (Tm/ha) en 2002 a 3.2 (Tm/ha) en 2016, lo que equivale a un incremento porcentual del 175.5 %. Las mejoras tanto en producción como en rendimiento se debe en su mayoría al impulso que ha dado el gobierno mediante

[1]El rendimiento de un producto i se determina como la producción en toneladas(Tm) del producto i sobre la superficie cosechada en hectáreas(ha) del producto i [28].

políticas públicas que han permitido el acceso a semilla mejorada, paquetes tecnológicos de alto rendimiento, entre otros.

Figura 1.1: Producción y rendimiento de maíz duro seco



Fuente: INEC
Elaboración: Autor

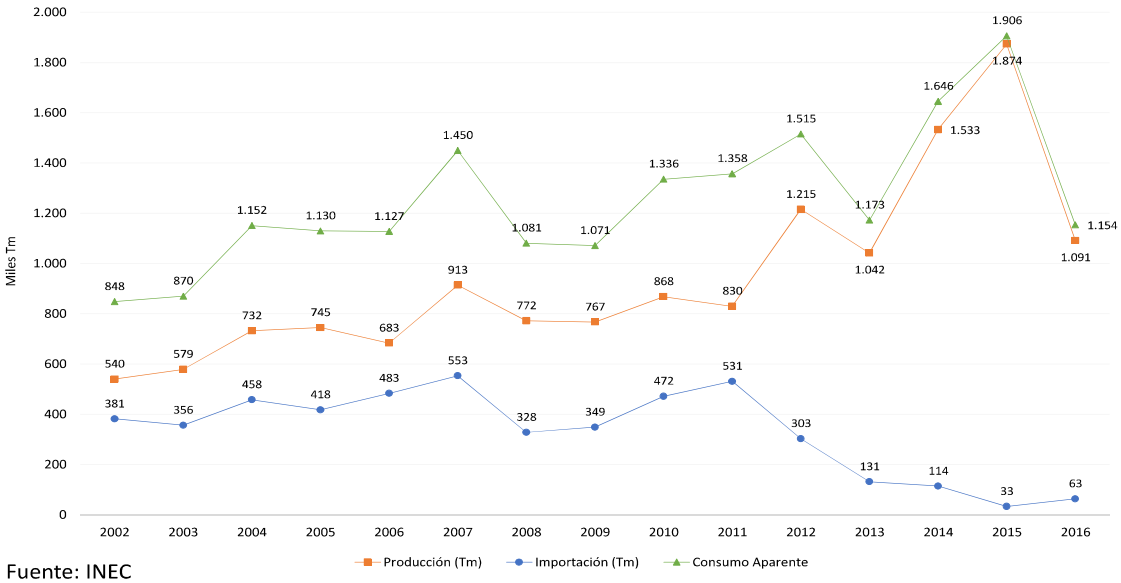
A partir del 2012 se evidencia un crecimiento sustancial en la producción de maíz duro seco en el país, haciendo cada vez más corta la brecha entre el consumo aparente nacional y la producción nacional, lo cual es un indicador de que a corto o largo plazo el Ecuador será un país autosuficiente de esta gramínea. En la figura 1.2 se observa tanto la producción, importación y consumo aparente nacional [2] [25]. Al 2016 la producción de maíz duro seco logró cubrir el 94,5 % del consumo aparente nacional, hecho acompañado de la reducción en las importaciones que para el mismo año fue de 63.398 Tm.

En la figura 1.3 se presentan las importaciones de maíz duro seco, se observa que a partir del 2011 se registra una tendencia decreciente en la importación de esta materia prima, esto se puede deber a factores como el incremento en la producción nacional y en el control que ejerce el Ministerio de Agricultura, Ganadería, Acuacultura y Pesca (MAGAP) en la asignación de cupos de importaciones donde las industrias que usan esta materia prima tienen el compromiso de comprar primero el 100 % de la producción nacional y registrar la compra en la Unidad de Registro de Transacciones Financieras (URTF) antes de solicitar el respectivo permiso de importación. Entre los principales países de los cuales se importa están Estados Unidos, Argentina, Brasil y Paraguay en ese orden de importancia, siendo es estos dos últimos años Estados Unidos y Argentina

[2]Consumo Interno Aparente (CA): El consumo interno aparente es la medida de la demanda nacional aparente de un país, es decir, es la producción más las importaciones menos las exportaciones.

los dos únicos proveedores de la gramínea.

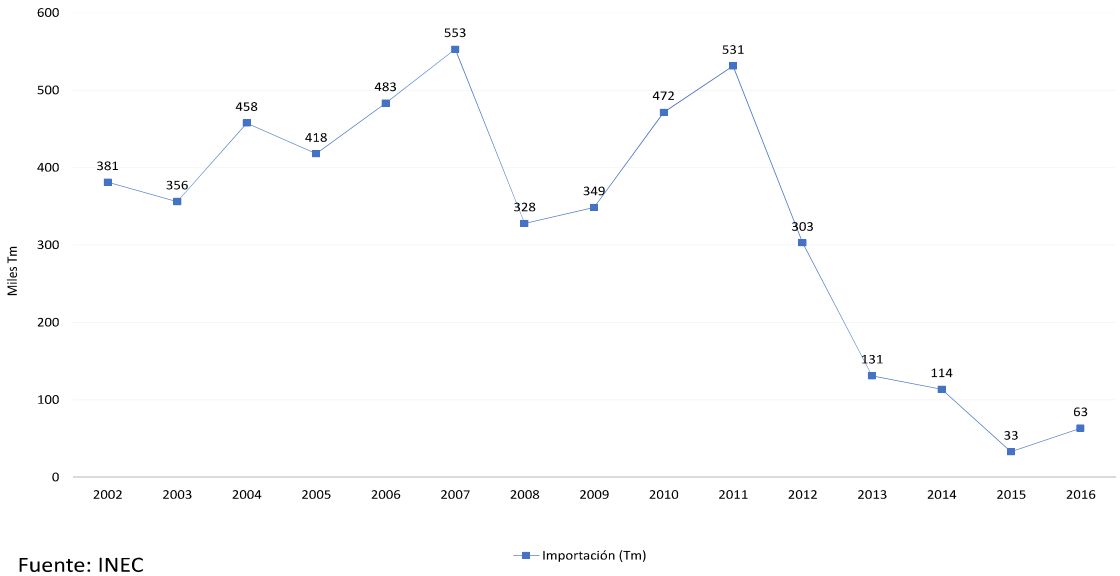
Figura 1.2: Consumo aparente nacional, producción nacional e importación



Fuente: INEC
Elaboración: Autor

En lo referente a las exportaciones de maíz duro seco del Ecuador, en todo el período histórico registrado ha presentado una tendencia a la baja, como puede observar en la figura 1.4, a partir del 2010 se registran los valores más bajos, con una tasa anual promedio de decrecimiento de exportaciones del 39,8 %. En los últimos cinco años los países destino de las exportaciones son Colombia, Bélgica, Estados Unidos, Panamá, Italia, Quatar y Suiza en ese orden de importancia. En los años comprendidos entre el 2002 y 2016 la

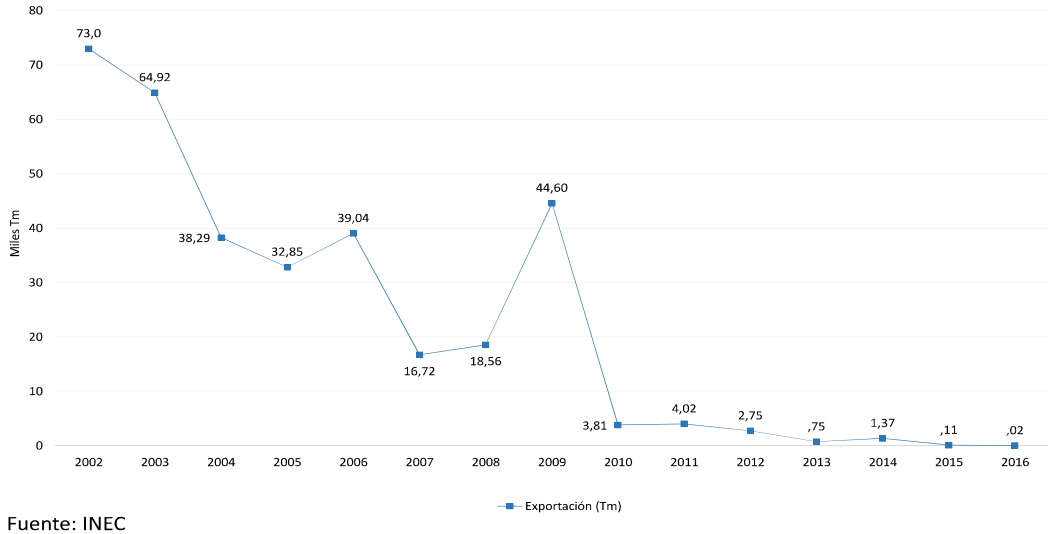
Figura 1.3: Importaciones de maíz duro seco



Fuente: INEC
Elaboración: Autor

producción aproximadamente se duplicó de 540 mil Tm a 1.09 millones de Tm, algo similar sucede con el rendimiento promedio que pasó de 1.8 (Tm/ha) a 3.2 (Tm/ha) a nivel nacional. La participación en la superficie sembrada para el 2016 corresponde a Los Ríos con 41 %, Manabí con 26 %, Guayas con 14 %, Loja con 9 %, Napo con 2 % y el resto de provincias con el 8 %. Debido al crecimiento de la población y de la industria aviar, porcina y ganadera el consumo de maíz duro seco creció en un 130 % en el período entre 2002 y 2016 [20].

Figura 1.4: Exportaciones de maíz duro seco



Fuente: INEC
Elaboración: Autor

1.3. Operativo de maíz duro seco

La Coordinación Nacional del Sistema de Información Nacional (CGSIN) del Ministerio de Agricultura, Ganadería, Acuacultura y Pesca (MAGAP) realiza el Operativo de Rendimientos Objetivos de Maíz Duro Seco a partir de la temporada de invierno del 2014, con el objetivo de determinar y actualizar el rendimiento nacional de maíz duro seco, así como sus condiciones de siembra e insumos.

Debido al éxito en la recolección de información y la calidad de los resultados, para el levantamiento de datos de la época de verano del 2014 se añadieron a la boleta de encuesta datos sobre la asociatividad, capacitación, método de siembra, acceso a agua y costos de producción por ser información valiosa y de interés para los análisis de la situación del cultivo [27].

El área para el levantamiento de las muestras fueron determinadas mediante imágenes

satelitales de alta resolución, que contienen la ubicación y el área de la superficie sembrada del cultivo de maíz duro seco a nivel nacional. Con los resultados obtenidos se estableció como área de investigación las provincias de Los Ríos, Guayas, Loja, Manabí, El Oro y Santa Elena por ser las provincias de mayor número de hectáreas sembradas de maíz duro seco [3].

Para el levantamiento de la información el cálculo del tamaño de la muestra se la realiza mediante la siguiente fórmula

$$n = \frac{Z^2 N p q}{e^2 (N-1) + Z^2 p q}, \tag{1.1}$$

donde Z : grados de confiabilidad (1,96), N : población, p : ocurrencia (0,5), q : no ocurrencia (0,5), e : error (0,05). Un ejemplo sobre la aplicación y obtención del número de la muestra se puede ver en el Boletín de Rendimiento de Maíz Duro Seco de Verano de 2014 [27].

La recolección de la información se la realiza en campo y en laboratorio. En el campo el investigador realiza la encuesta al agricultor según la Boleta de Recolección (ver Anexo 1), luego se recolecta los datos de la siembra que serán usados en el laboratorio [27].

Las principales variables que se investigan son:

- **Superficie sembrada:** área total en hectáreas (ha) dedicadas a la siembra de maíz duro seco.
- **Fertilización:** cantidad de quintales (qq, 50 kg) utilizados por hectárea de los principales fertilizantes medidos según sus macronutrientes, nitrógeno, fósforo y potasio.
- **Cantidad de material vegetativo:** se registra la cantidad de semilla usada por hectárea, medido en kg/ha; se consideran valores válidos los que se encuentran entre $13kg/ha$ y $30kg/ha$, estos valores serán usados en el tratamiento de la base de datos.
- **Capacitación recibida:** se registra si en el último año el productor ha recibido cursos o programas de capacitación.
- **Gasto en producción:** se registra el costo total del cultivo por hectárea durante el ciclo, debe considerarse la preparación del suelo, la semilla, fertilizantes, agroquímicos, cosecha, mano de obra contratada, mano de obra propia y maquinaria.
- **Acceso a fuente de agua:** si el agricultor tiene acceso o no a una fuente de agua, que puede ser ríos, lagunas, represas, etc.

- **Fecha de siembra y cosecha:** se registra día, mes y año de siembra y la fecha programada de cosecha; la fecha de siembra será considerada para asignar la temperatura promedio del ciclo y la precipitación acumulada del ciclo, que son datos obtenidos del INAMHI. La fecha de cosecha debe sr mayor que la de siembra y no mayor a un año, datos considerados en el tratamiento del la base.
- **Plan semilla:** si el agricultor utilizó el kit del Plan Semilla de alto Rendimiento.
- **Rendimiento:** es el rendimiento en toneladas métricas por hectárea (Tm/ha). Se calcula de acuerdo a la fórmula determinada en el Operativo [3].
- **Número de plantas:** se registran el número de plantas existentes en 10 metros lineales, tal como se describe en el Manual del Investigador. Debe registrar un mínimo de 7 y máximo de 30 plantas [6].
- **Número de mazorcas:** se registra el número de mazorcas existente en 10 metros lineales. Usualmente igual al número de plantas.
- **Número de plantas por hectárea:** se registra el número de plantas en una hectárea de siembra, con un mínimo de 40 000 $plantas/ha$ y un máximo de 60 000 $plantas/ha$.
- **Distancia entre surco:** se mide la distancia entre surcos (cm), deberá ser mínimo 50 cm y máximo 150 cm [6].
- **Miembro de asociación:** registra si el agricultor es miembro de una asociación.
- **Cosecha mecanizada:** registra si la cosecha fue realizada de forma mecanizada o manual.

Los campos de la temperatura promedio y precipitación acumulada se obtuvieron de las estaciones del INAMHI y se asignaron de acuerdo a la ubicación geográfica y fecha de siembra de cada cultivo, considerando que el ciclo de cultivo en Los Ríos, Manabí, Guayas, Santa Elena y El Oro es de cuatro meses aproximadamente y en zonas altas como Loja el ciclo es de cinco meses aproximadamente [7], [30].

Capítulo 2

Marco teórico

En este capítulo se presentan los fundamentos teóricos para el desarrollo de modelos de regresión lineal múltiple, así como los supuestos y las pruebas de validación de los resultados. En la segunda parte se describen las redes neuronales y se presentan los conceptos y notaciones usadas para el ajuste de los pesos sinápticos mediante el algoritmo backpropagation y sus versiones mejoradas.

2.1. Regresión lineal múltiple

2.1.1. Modelo de Regresión Múltiple

El modelo general de regresión lineal múltiple es usado para estudiar la relación entre una variable dependiente y una o más variables independientes. Matemáticamente puede expresarse como

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u, \tag{2.1}$$

donde y es la variable dependiente o explicada, $x_1, x_2, \ldots, x_k$ son k variables independientes o explicativas. La ecuación (2.1) es conocida como la ecuación de regresión lineal poblacional de y sobre $x_1, x_2, \ldots, x_k$. El término u es conocido como el término de error o perturbación, β_0 es el intercepto y cada β_i es el parámetro asociado a x_i, donde $i = 1, \ldots, k$.

Es importante considerar que el término *lineal* en un modelo de regresión lineal múltiple significa que la ecuación (2.1) es lineal en los *parámetros*. Independientemente del número de variables explicativas que se incluyan en el modelo siempre existirán otras dis-

tintas de $x_1, x_2, \ldots, x_k$ que no se pueden incluir en el modelo ya sea por desconocimiento y por falta de valores observados de los mismos; también hay que considerar que las variables usadas en el modelo están sujetas a errores de medida. Todas estas variables están contenidas en u. Con estas consideraciones se presenta el primer supuesto del modelo de regresión lineal múltiple.

***Supuesto* 2.1** (Linealidad)**.** *El modelo de regresión poblacional es lineal en los parámetros y se expresa como la ecuación 2.1*

A diferencia de lo que se podría creer, la relación entre las variables independientes y la variable dependiente puede no ser lineal, ya que el supuesto de linealidad solamente hace referencia a los parámetros del modelo. Un ejemplo es el *modelo de elasticidad* que emplea logaritmos tanto para la variable dependiente como para las variables independientes.

***Supuesto* 2.2.** *Dado los valores de las variables independientes, la media condicional de u es cero. Es decir*

$$E\left(u|x_1, x_2, \ldots, x_k\right) = 0. \tag{2.2}$$

El supuesto 2 nos permite establecer que la regresión de y respecto de las variables independientes $x_1, x_2, \ldots, x_k$ es la media condicional $E\left(y|x_1, x_2, \ldots, x_k\right)$. En efecto, si tomamos esperanzas condicionales en ambos lados de la ecuación (2.1) tenemos

$$\begin{aligned} E\left(y|x_1, x_2, \ldots, x_k\right) &= E\left(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u|x_1, x_2, \ldots, x_k\right) \\ &= E\left(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k|x_1, x_2, \ldots, x_k\right) + E\left(u|x_1, x_2, \ldots, x_k\right) \\ &= \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k. \end{aligned} \tag{2.3}$$

Se entenderá entonces que la función de regresión poblacional nos dice cómo varía el valor promedio de y respecto de la variación de las variables $x_1, x_2, \ldots, x_k$. Dado que en la práctica se dispone de una muestra de los valores de y para valores fijos de x_i $(i = 1, \ldots, k)$, el objetivo será estimar la función de regresión poblacional a partir de una muestra aleatoria tomada de la población.

Los siguientes supuestos serán usados más adelante para obtener resultados importantes.

***Supuesto* 2.3.** *Se tiene una muestra aleatoria de n observaciones de una población que sigue el modelo descrito por la ecuación (2.1).*

***Supuesto* 2.4.** *No existe colinealidad perfecta, es decir, ninguna de las variables independientes es combinación lineal de las otras y tampoco es constante.*

***Supuesto* 2.5** (Homocedasticidad)**.** *La varianza del residual* $\mathbf{u}$ *dado cualquier valor de las variables independientes es constante, es decir,*

$$V(\mathbf{u}|\mathbf{X}) = \sigma^2 \mathbf{I_n}, \tag{2.4}$$

donde σ^2 *es una constante e* $\mathbf{I_n}$ *es la matriz identidad de orden* n*.*

***Supuesto* 2.6** (Normalidad)**.** *El error* u *es independiente de las variables explicativas y sigue una distribución normal de media cero y varianza constante, es decir*

$$u \sim N\left(0, \sigma^2\right). \tag{2.5}$$

En forma matricial, tenemos

$$\mathbf{u} \sim N\left(\mathbf{0_n}, \sigma^2 \mathbf{I_n}\right), \tag{2.6}$$

donde $\mathbf{0_n}$ *es el vector de medias e* $\mathbf{I_n}$ *es la matriz identidad de orden* n*.*

Los supuestos del 1 al 6 son conocidos como supuestos del modelo de regresión lineal clásico.

2.1.2. Estimación

La estimación de la función de regresión poblacional, a partir de una muestra de n observaciones, es la función de regresión muestral que se define como

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_k x_k, \tag{2.7}$$

donde $\hat{y}$ es el estimador muestral de y y cada $\hat{\beta}_i$ es el estimador muestral de β_i, $i = 0, \ldots, k$.

Se debe estimar la función de regresión poblacional a partir de la función de regresión muestral, es decir, se estimarán los parámetros β_i, $i = 1, \ldots, k$.

De la función de regresión muestral podemos determinar el error de la siguiente manera

$$u = y - \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k. \tag{2.8}$$

A partir de los valores estimados $\hat{\beta}_i$ podemos obtener el **residual**, que es la estimación de u:

$$\hat{u} = y - \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_k x_k, \tag{2.9}$$

Para estimar los parámetros del modelo de regresión lineal múltiple se usará el método de mínimos cuadrados ordinarios (MCO), que determina la obtención de los estimadores de los parámetros de tal manera que la suma de los residuales al cuadrado sea mínima,es decir, la expresión

$$\sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} \left(y_i - \left(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \ldots + \hat{\beta}_k x_{ik} \right) \right)^2 \tag{2.10}$$

debe ser lo más pequeña posible.

Empleando derivadas parciales respecto de cada uno de los parámetros e igualando a cero las ecuaciones obtenidas tenemos

$$\sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \ldots - \hat{\beta}_k x_{ik} \right) = 0, \tag{2.11}$$

$$\sum_{i=1}^{n} x_{ij} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \ldots - \hat{\beta}_k x_{ik} \right) = 0, \quad j = 1, \ldots, k \tag{2.12}$$

Estas $k+1$ ecuaciones son conocidas como condiciones de primer orden de MCO.

Sean los vectores columna $\hat{\beta} = [\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k]'$, $\hat{\mathbf{u}} = [\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_n]'$, $\mathbf{y} = [y_1, y_2, \ldots, y_n]'$ y los vectores columna $\mathbf{x_j} = [x_{1j}, x_{1j}, \ldots, x_{nj}]'$, $j = 1, \ldots, k+1$ que forman la matriz $\mathbf{X}$ (el vector $\mathbf{x_1}$ es un vector de unos). Con esta notación podemos expresar de manera matricial las condiciones de primer orden:

$$\begin{aligned} \sum_{i=1}^{n} \hat{u}_i^2 &= \hat{\mathbf{u}}'\hat{\mathbf{u}} \\ &= (\mathbf{y} - \mathbf{X}\hat{u})' (\mathbf{y} - \mathbf{X}\hat{u}) \\ &= \mathbf{y}'\mathbf{y} - 2\hat{\beta}'\mathbf{X}'\mathbf{y} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} \end{aligned} \tag{2.13}$$

Tomando la derivada de la expresión anterior e igualando a cero

$$\begin{aligned} \frac{\partial}{\partial \hat{\beta}} \left(\mathbf{y}'\mathbf{y} - 2\hat{\beta}'\mathbf{X}'\mathbf{y} + \hat{\beta}'\mathbf{X}'\mathbf{X}\hat{\beta} \right) &= 0 \\ -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\beta} &= 0 \\ \mathbf{X}'\mathbf{X}\hat{\beta} &= \mathbf{X}'\mathbf{y} \end{aligned} \tag{2.14}$$

Por lo tanto, el estimador MCO se pueden expresar de forma matricial de la siguiente

manera:

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{2.15}$$

Analizando la expresión (2.15), $\hat{\beta}$ existirá siempre y cuando $(\mathbf{X}'\mathbf{X})^{-1}$ exista. Esto se garantiza con el supuesto 3, que implica que la matriz $\mathbf{X}$ es de rango completo, y por lo tanto su transpuesta $\mathbf{X}'$ también tiene rango completo. Por lo tanto el producto $\mathbf{X}'\mathbf{X}$ es una matriz cuadrada con rango completo y definida positiva, lo que garantiza la existencia de su inversa $(\mathbf{X}'\mathbf{X})^{-1}$.

Con los supuestos hasta ahora mencionados podemos determinar que estadísticamente el estimador $\hat{\beta}$ sigue una distribución normal con los parámetros que se describen en el siguiente teorema.

***Teorema* 2.1.** *El estimador de mínimos cuadrados $\hat{\beta}$ del vector de parámetros β tiene una distribución normal multivariante con vector de medias β y matriz de varianzas y covarianzas $\sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}$, notado como*

$$\hat{\beta} \sim N\left(\beta, \sigma_u^2(\mathbf{X}'\mathbf{X})^{-1}\right) \tag{2.16}$$

Demostración. El supuesto 2 en forma matricial se puede escribir de la siguiente manera

$$E(\mathbf{u}|\mathbf{X}) = \mathbf{0}.$$

Para cada u_i se tiene que $E(u_i) = E_X(E(u_i|\mathbf{X})) = E_X(0) = 0$, por lo tanto $E(\mathbf{u}) = \mathbf{0}$.

Ahora, de la ecuación (2.15) reemplazando el valor de $\mathbf{y}$ tenemos

$$\begin{aligned}\hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mathbf{u}) \\ &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}. \end{aligned} \tag{2.17}$$

Tomando esperanzas en (2.17):

$$\begin{aligned} E\left(\hat{\beta}\right) &= E\left(\beta + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{u}\right) \\ &= \beta + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'E(\mathbf{u}) \\ &= \beta. \end{aligned} \tag{2.18}$$

Para la parte de la varianza, la ecuación (2.15) podemos escribirla como

$$\begin{aligned} \hat{\beta} &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} \\ &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'(\mathbf{X}\beta + \mathbf{u}) \\ &= \beta + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{u} \\ \hat{\beta} - \beta &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{u} \end{aligned} \tag{2.19}$$

Dado el supuesto $V(\mathbf{u}) = E(\mathbf{u}\mathbf{u}') = \sigma^2\mathbf{I}$:

$$\begin{aligned} V\left(\hat{\beta}\right) &= E\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right] \\ &= E\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right] \\ &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'E\left[\mathbf{u}\mathbf{u}'\right]\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \\ &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\sigma^2\mathbf{I}\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \\ &= \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \\ &= \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1} \end{aligned} \tag{2.20}$$

□

Definamos qué es un estimador insesgado y eficiente, ya que estos términos serán usados en la demostración del teorema de *Gauss-Markov*, el cual nos garantiza que el estimador hallado por el método MCO es un estimador insesgado y de mínima varianza, es decir, es el mejor estimador lineal insesgado.

***Definición* 2.1.** *Un estimador $\hat{\beta}_i$ del parámetro β_i es insesgado si su esperanza matemática coincide con el verdadero parámetro β_i, es decir, $E(\hat{\beta}_i) = \beta_i$. En el caso multidimensional, un vector de estimadores $\hat{\beta}$ es insesgado si $E(\hat{\beta}) = \beta$.*

En la primera parte del teorema (2.1) se demuestra que el estimador $\hat{\beta}$ es insesgado.

***Definición* 2.2.** *Un estimador insesgado $\hat{\beta}_i$ es más eficiente que otro estimador $\tilde{\beta}_i$ también insesgado, si la varianza muestral de $\hat{\beta}_i$ es menor que la de $\tilde{\beta}_i$, es decir, $V(\hat{\beta}_i) <$*

$V(\tilde{\beta}_i)$. *En el caso multidimensional, un vector de estimadores insesgados* $\hat{\beta}$ *es más eficiente que otro* $\tilde{\beta}$*, si la diferencia entre las matrices de varianzas y covarianzas* $V(\hat{\beta}) - V(\tilde{\beta})$ *es una matriz definida negativa.*

***Teorema* 2.2** (Gauss-Markov)**.** *Bajo los supuestos básicos del modelo clásico, el estimador de mínimos cuadrados* $\hat{\beta}$ *es el más eficiente en la clase de estimadores lineales e insesgados de* β*.*

Demostración. Sea $\tilde{\beta}$ un estimador lineal e insesgado de β. Por tanto, $\tilde{\beta}$ puede escribirse de la forma

$$\tilde{\beta} = \mathbf{C}\mathbf{y}, \tag{2.21}$$

donde $\mathbf{C}$ es una matriz de orden $k+1 \times n$ de números fijos. Y además

$$\begin{aligned} E(\tilde{\beta}) &= E(\mathbf{C}\mathbf{y}) \\ &= E(\mathbf{C}\mathbf{X}\beta + \mathbf{C}\varepsilon) \\ &= \mathbf{C}\mathbf{X}\beta \\ &= \beta, \end{aligned} \tag{2.22}$$

lo que implica que $\mathbf{C}\mathbf{X} = \mathbf{I}$. La amtriz de varianzas y covarianzas de $\tilde{\beta}$ es

$$\begin{aligned} V(\tilde{\beta}) &= E\left[(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'\right] \\ &= \mathbf{C}E[\varepsilon\varepsilon']\mathbf{C}' \\ &= \sigma^2\mathbf{C}\mathbf{C}' \end{aligned} \tag{2.23}$$

Sea $\mathbf{D}$ tal que

$$\mathbf{D} = \mathbf{C} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}', \tag{2.24}$$

entonces $\mathbf{D}\mathbf{X} = \mathbf{0}$ porque $\mathbf{C}\mathbf{X} = \mathbf{I}$. De modo que

$$\begin{aligned} \mathbf{C}\mathbf{C}' &= \left[\mathbf{D} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]\left[\mathbf{D}' + \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right] \\ &= \mathbf{D}\mathbf{D}' + (\mathbf{X}'\mathbf{X})^{-1} \end{aligned} \tag{2.25}$$

Sustituyendo $\mathbf{CC}'$ en (2.23) tenemos

$$V(\tilde{\beta}) = \sigma^2\mathbf{DD}' + \sigma^2\mathbf{X}'\mathbf{X}^{-1} \tag{2.26}$$

$$V(\tilde{\beta}) - V(\hat{\beta}) = \sigma^2\mathbf{DD}', \tag{2.27}$$

donde $\mathbf{DD}'$ es una matriz semidifinida positiva porque para cualquier vector $\mathbf{q}$ de dimensión adecuada se tiene que $\mathbf{q}'\mathbf{DD}'\mathbf{q} = \eta'\eta \geq 0$. Por lo tanto la diferencia de las dos matrices de varianzas y covarianzas es una matriz semidefinida positiva, lo cual concluye la demostración. □

2.1.3. Validación

Significancia de la regresión

Se presentan algunas definiciones básicas usadas en la validación del modelo de regresión.

***Definición* 2.3.**

- *Se define la suma de cuadrados total (**SCT**) del modelo mediante la siguiente fórmula*

$$SCT = \sum_{i=1}^{n} (y_i - \bar{y})^2, \tag{2.28}$$

donde $\bar{y}$ es el promedio de la variable explicada.

- *Se define la suma de cuadrados residuales (**SCR**) de la siguiente manera*

$$SCR = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \hat{u}_i^2, \tag{2.29}$$

- *y la suma de cuadrados explicada (**SCE**) mediante la fórmula*

$$SCE = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2. \tag{2.30}$$

La SCT mide la variación muestral total de las y_i, es decir, qué tan dispersos están las y_i en la muestra. La SCR mide la variación de los $\hat{u}_i$ y la SCE mide la variación muestral de las las $\hat{y}_i$. Podemos descomponer la SCT en dos partes, la variación explicada (SCE)

y la variación sin explicar (SCR):

$$\begin{aligned}\sum_{i=1}^{n}(y_i-\overline{y})^2 &= \sum_{i=1}^{n}\left[(y_i-\hat{y}_i)+(\hat{y}_i-\overline{y})\right]^2\\ &= \sum_{i=1}^{n}\left[\hat{u}_i+(\hat{y}_i-\overline{y})\right]^2\\ &= \sum_{i=1}^{n}\hat{u}_i^2+2\sum_{i=1}^{n}\hat{u}_i(\hat{y}_i-\overline{y})+\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2\\ &= SCR+2\sum_{i=1}^{n}\hat{u}_i(\hat{y}_i-\overline{y})+SCE\\ &= SCR+SCE, \end{aligned} \tag{2.31}$$

se tiene finalmente

$$SCT = SCR + SCE. \tag{2.32}$$

Con este resultado podemos definir la R *cuadrada* de la regresión, que será tratada más adelante en la bondad de ajuste.

***Definición* 2.4.** *La R cuadrada de la regresión (o coeficiente de determinación), se define como*

$$\begin{aligned}R^2 &= \frac{SCE}{SCT}\\ &= 1-\frac{SCR}{SCT}. \end{aligned} \tag{2.33}$$

Se interpreta como la proporción de la variación muestral de la variable dependiente y que es explicada por las variables independientes; es usada como estadístico para medir la bondad de ajuste de un modelo de regresión [49].

La prueba de significancia del modelo de regresión nos permite determinar si hay una relación lineal entre la variable dependiente y y cualquiera de las variables independientes $x_1,\ldots,x_k$. Este procedimiento se considera como una prueba global del modelo. Tenemos como hipótesis nula contra la hipótesis alternativa

$$H_0: \quad \beta_1 = \cdots = \beta_k = 0 \tag{2.34}$$

$$H_1: \quad \beta_j \neq 0, \text{para al menos una } j, 1 \leq j \leq k. \tag{2.35}$$

El rechazo de la hipótesis nula implica que al menos una de las variables independien-

tes contribuye al modelo en forma significativa. Para realizar esta prueba se define el estadístico F de la siguiente manera:

***Definición* 2.5.** *El estadístico F o cociente F se define como*

$$F = \frac{SCE/k}{SCR/n-k-1}. \tag{2.36}$$

Se puede demostrar que bajo el supuesto (2.34) F está distribuida como una variable aleatoria F con $(k, n-k-1)$ grados de libertad, notado $F \sim F_{k,n-k-1}$.

***Teorema* 2.3.** *Bajo los supuestos del modelo lineal clásico y la hipótesis nula dada en (2.34), el estadístico F sigue una distribución $F_{k,n-k-1}$.*

Demostración. El siguiente resultado está demostrado en [29] y hace uso del supuesto de normalidad de u: SCE y SCR son independientes y

$$\frac{SCE}{\sigma^2} \sim \chi^2_k \tag{2.37}$$

$$\frac{SCR}{\sigma^2} \sim \chi^2_{n-k-1}. \tag{2.38}$$

Con este resultado y por definición de la distribución F se tiene que

$$\frac{\frac{SCE}{k\sigma^2}}{\frac{SCR}{\sigma^2(n-k-1)}} = \frac{\frac{SCE}{k}}{\frac{SCR}{(n-k-1)}} \sim F_{k,n-k-1}. \tag{2.39}$$

□

La hipótesis nula H_0 se rechazará a favor de H_1 cuando F sea lo suficientemente grande, y esto depende del nivel de significancia elegido. En el presente trabajo se ha elegido el 5 % como nivel de significancia. Sea c el percentil 95 de la distribución $F_{k,n-k-1}$, entonces H_0 se rechaza a favor de H_1 al nivel de confianza determinado si $F > c$.

Cuando rechazamos H_0 decimos que las variables independientes son estadísticamente significativas conjuntamente al nivel de significancia especificado. Pero en sí, esta prueba no me permite decir cual o cuales variables influyen en la variable dependiente y, ya que puede darse el caso de que todas las variables tengan un efecto sobre y o solo algunas.

Una vez determinado que al menos uno de los regresores es diferente de cero, es importante saber cuál de ellos influye en el modelo ya que si agregamos una variable al

modelo de regresión la suma de cuadrados de la regresión aumenta y la suma de cuadrados residuales disminuye. Las hipótesis para probar la significancia de cualquier coeficiente individual, por ejemplo β_j, son:

$$H_0: \quad \beta_j = 0 \tag{2.40}$$

$$H_1: \quad \beta_j \neq 0. \tag{2.41}$$

Para realizar la prueba de hipótesis se necesitan los siguientes resultados.

***Teorema* 2.4.** *Bajo los primeros 5 supuestos del modelo de regresión clásico se tiene que*

$$V(\hat{\beta}_j) = \frac{\sigma^2}{SCT_j(1 - R_j^2)}, \tag{2.42}$$

para $1 \leq j \leq k$, *donde* $SCT_j = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$ *es la variación muestral total en* x_j *y* R_j^2 *es la* R *cuadrada de la regresión de* x_j *sobre todas las otras variables independientes incluyendo un intercepto.*

Demostración. La demostración de este teorema se encuentra en [49], teorema *3.2.* □

***Teorema* 2.5.** *Bajo los supuestos del modelo de regresión clásico, se tiene que*

$$\hat{\beta}_j \sim N\left(\beta_j, V(\hat{\beta}_j)\right), \tag{2.43}$$

donde $V(\beta_j) = \sigma^2 c_{ii}$ *con* c_{ii} *el elemento diagonal de la matriz* $(\mathbf{X}\mathbf{X}')^{-1}$ *que le corresponde a* β_j.

Demostración. A partir de la ecuación (2.17) cada $\hat{\beta}_j$, $1 \leq j \leq k$, se puede expresar como

$$\hat{\beta}_j = \beta + \sum_{i=1}^{n} w_{ij} u_i, \tag{2.44}$$

donde w_{ij} es determinístico pues depende de las variables independientes x_i. Entonces $\hat{\beta}_j$ es solo un combinación lineal de los errores u_i, y puesto que éstos están normalmente distribuidos se concluye que $\hat{\beta}_j$ sigue una distribución normal. Por el supuesto de normalidad se concluye que $E(\hat{\beta}_j) = \beta_j$ y $V(\hat{\beta}_j)$ como se dedujo en el teorema 2.4. □

Con el resultado del teorema (2.5) podemos determinar la ***desviación estándar*** de

$\hat{\beta}_j$, que no es más que la raíz cuadrada de la varianza

$$de(\hat{\beta}_j) = \frac{\sigma}{\left[SCT_j\left(1-R_j^2\right)\right]^{1/2}}. \tag{2.45}$$

Dado que no se conoce σ, entonces se sustituye por su estimador $\hat{\sigma}$.Esto se define como el ***error estándar*** de $\hat{\beta}_j$:

$$ee(\hat{\beta}_j) = \frac{\hat{\sigma}}{\left[SCT_j\left(1-R_j^2\right)\right]^{1/2}}. \tag{2.46}$$

Además se tiene que

$$\frac{(\hat{\beta}_j - \beta_j)}{de(\hat{\beta}_j)} \sim N(0,1). \tag{2.47}$$

***Teorema* 2.6.** *Bajo los supuestos del modelo de regresión clásico,*

$$t = \frac{\hat{\beta}_j - \beta_j}{ee(\hat{\beta}_j)} \sim t_{n-k-1}. \tag{2.48}$$

Demostración. Dado que la variable z definida por

$$z = \frac{Z_1}{\sqrt{\dfrac{Z_2}{m}}}, \tag{2.49}$$

sigue una distribución t de Student con m grados de libertad, donde $Z_1 \sim N(0,1)$ y $Z_2 \sim \chi^2_m$, si tomamos como Z_1 el resultado obtenido en (2.47) y el siguiente resultado

$$\frac{(n-k-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-k-1}, \tag{2.50}$$

obtenemos que

$$\frac{\dfrac{(\hat{\beta}_j - \beta_j)}{\sigma}}{\sqrt{\dfrac{(n-k-1)\hat{\sigma}^2}{\sigma^2}/(n-k-1)}} = \frac{(\hat{\beta}_j - \beta_j)}{\hat{\sigma}/\left[SCT_j\left(1-R_j^2\right)\right]^{1/2}} \tag{2.51}$$

$$= \frac{\hat{\beta}_j - \beta_j}{ee(\hat{\beta}_j)}. \tag{2.52}$$

sigue una distribución t_{n-k-1}. □

Con estadístico de prueba dado por (2.48) se rechaza la hipótesis nula $H_0 : \beta_j = 0$ si $|t| > t_{\alpha/2,n-k-1}$, donde α es el nivel de significancia elegido, en este caso del 5 %. Cuando se rechaza H_0 en favor de H_1 al nivel 5 % se suele decir que x_j es *estadísticamente significativa* al nivel 5 %.

Multicolinealidad

Se dice que existe *multicolinealidad* cuando existe una relación lineal entre algunas o todas las variables explicativas de un modelo de regresión. Según [12] se presentan dos tipos de multicolinealidad: la ***multicolinealidad perfecta*** y la ***multicolinealidad menos que perfecta***. La multicolinealidad perfecta se presenta cuando existe una relación lineal exacta, es decir, cuando se satisface la condición:

$$\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_k x_k = 0, \tag{2.53}$$

donde $\lambda_1, \ldots, \lambda_k$ son constantes tales que no todas son simultáneamente iguales a cero. La multicolinealidad menos que perfecta existe cuando se satisface la condición:

$$\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_k x_k + \epsilon = 0, \tag{2.54}$$

donde ϵ es un término estocástico. Notar que si existe colinealidad perfecta entonces no existe la matriz $(\mathbf{X}'\mathbf{X})^{-1}$ y por lo tanto no se podrá calcular $\hat{\beta}$. Entre las consecuencias que se presentan debido a la multicolinealidad tenemos:

- *Los estimadores MCO presentan varianzas y covarianzas que crecen y dificultan la estimación*
  Como se presenta la varianza en (2.42), a medida que R_j^2 tiende a uno, es decir, mientras aumenta la colinealidad, la expresión $\dfrac{1}{(1-R_j^2)}$ tiende a infinito, y la varianza del mismo modo.
  La expresión $\dfrac{1}{(1-R_j^2)}$ se conoce como ***factor inflación de la varianza (FIV)***.
  Una consecuencia directa del incremento de R_j^2 es que los valores del estadístico t disminuyan y por ende se acepta con mayor facilidad la hipótesis nula.

- *Se presenta una R^2 alta aunque pocas variables sean significativas*
  Debido a que de manera individual uno o más coeficientes pueden no ser significa-

tivos, aún así, R^2 puede ser alto y con base al estadístico F se puede rechazar la hipótesis nula (2.34).

- *Los estimadores de MCO son muy sensibles a cambios pequeños en los datos*
  Siempre que no exista multicolinealidad perfecta, se puede calcular el valor de los estimadores, sin embrago, las estimaciones y sus errores se tornan muy sensibles a cambios pequeños en los datos.

Existen varias técnicas para detectar la multicolinealidad, entre ellas, se busca cuales reflejen mejor el grado de multicolinealidad. A continuación se presentan las técnicas más usuales al momento de determinar el grado de multicolinealidad:

- *Matriz de correlación*
  La presencia de correlaciones altas entre pares de variables es un indicio de presencia de multicolinealidad. Pero, en caso de existir multicolinealidad entre una variable y el resto, esta no se verá reflejada en la matriz aún cuando la correlación en parejas sean bajas.
  Se define la matriz de correlaciones como

$$C = \begin{pmatrix} 1 & c_{12} & c_{13} & \dots & c_{1k} \\ c_{21} & 1 & c_{22} & \dots & c_{2k} \\ c_{31} & c_{32} & 1 & \dots & c_{3k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{k1} & c_{k2} & c_{k3} & \dots & 1 \end{pmatrix}, \tag{2.55}$$

  cada entrada c_{ij} se define por

$$c_{ij} = corr(x_i, x_j) = \frac{s_{ij}}{s_i s_j}, \text{ con } 1 \leq i, j \leq\leq k, \tag{2.56}$$

  donde s_{ij} es la covarianza muestral entre x_i y x_j; s_i y s_j son las desviaciones estándar de las variables x_i y x_j respectivamente. El valor de las correlaciones se mantienen entre -1 y 1: $-1 \leq r_{ij} \leq 1$, por lo tanto si se observan valores altos de correlación se tendría indicios de multicolinealidad.

- *Factor de inflación de varianza*
  El factor inflación de la varianza FIV:

$$FIV = \frac{1}{(1 - R_j^2)}, \tag{2.57}$$

nos sirve para medir el efecto que produce la multicilinealidad en el incremento de la varianza. En la práctica, si el FIV de una variable es superior a 10 (cuando R_j^2 es mayor que 0,90) se dice que la esa variable es muy colineal.

Cuando se presenta multicolinelaidad se recomienda realizar ciertas técnicas para corregir este problema

- *Regresión por componentes principales*
  El método de componentes principales combate la multicolinealidad al definir un nuevo conjunto de regresores llamados componentes principales, tales regresores deben ser ortogonales, es decir, no deben presentar multicolinealidad. Detrás de este análisis por factores existen varias técnicas y teoría vasta que queda fuera de este estudio. Para información detallada se puede consultar [44],[35].

- *Reespecificación del modelo*
  Cuando se tiene información *a priori* del modelo estudiado se puede redefinir los regresores que presentan colinealidad mediante alguna función que preserve la información de las variables originales pero que reduzca la multicolinealidad.

- *Eliminación de variables*
  Una de las soluciones más simples para enfrentar la multicolinealidad es eliminar la o las variables colineales y que sabemos aportan la misma información al modelo. Para eliminar se analiza la información obtenida de las técnicas que detectan el grado de multicolinealidad.

Heterocedasticidad

Cuando el supuesto de homocedasticidad no se cumple, es decir, cuando la varianza del residual dado cualquier valor de las variables independientes no es constante, decimos que el modelo presenta heterocedasticidad.
Según [12] y [33] las causas pueden ser:

- Por la presencia de datos atípicos o aberrantes, es decir, datos que son muy pequeños o muy grandes en relación con las demás observaciones de la muestra.

- Cuando el modelo no está correctamente especificado, esto sucede cuando se omiten variables importantes.

- Cuando los datos provienen de promedios de los distintos grupos en que se han dividido los individuos encuestados.

En presencia de heterocedasticidad los estimadores de las varianzas, $V(\hat{\beta}_j)$, son sesgados y como los errores estándar están directamente relacionados con estas varianzas entonces dejan de ser válidos para la construcción de intervalos de confianza y de estadísticos t. La presencia de heterocedasticidad implica que los estadísticos t no tienen distribuciones t de Student al igual que los estadísticos F dejan de seguir una distribución F. El estimador de MCO deja de sel el de menor varianza así como establece el teorema de *Gauss-Markov*.

Examinaremos algunos métodos para la detectar heterocedasticidad.

- ***Método gráfico***

  El método gráfico se realiza la regresión bajo el supuesto de homocedasticidad, luego se hace uso de los residuos muestrales elevados al cuadrado, $\hat{u}_i^2$, los mismos que son graficados frente a los valores de $\hat{y}_i$. El objetivo es ver si el valor medio estimado de y está relacionado con el residuo al cuadrado. Si bien es un método informal, sirve se guía en las técnicas de corrección que se basan en transformaciones de la variable independiente. Se esperaría que los gráficos no presenten relaciones lineales ni cuadráticas sino puntos aleatorios sin tendencia.

- ***Prueba de Breusch-Pagan***

  Consideremos como hipótesis nula el supuesto de homocedasticidad

$$H_0 : V(u|x_1, \ldots, x_k) = \sigma^2. \tag{2.58}$$

  Por el supuesto 2 se tiene que $V(u|x) = E(u^2|x)$, de esta manera la hipótesis nula es equivalente a

$$H_0 : E(u^2|x_1, \ldots, x_k) = E(u^2) = \sigma^2. \tag{2.59}$$

  Este supuesto no se cumplirá cuando u^2 esté relacionado con las variables explicativas. Suponiendo una relación lineal se tiene

$$u^2 = \delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k + v, \tag{2.60}$$

  donde v es un término de error. Dado que la variable dependiente es u^2, la hipótesis nula es

$$H_0 : \delta_0 = \delta_1 = \cdots = \delta_k. \tag{2.61}$$

Como se hizo anteriormente, el estadístico F nos servirá para determinar la significancia general en esta hipótesis. En términos muestrales la ecuación (2.60) se expresa como

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \cdots + \delta_k x_k + \hat{v}. \tag{2.62}$$

El estadístico F usado en la prueba es

$$F = \frac{R^2_{\hat{u}^2}/k}{(1 - R^2_{\hat{u}^2})/(n-k-1)}, \tag{2.63}$$

donde $R^2_{\hat{u}^2}$ es la R cuadrada de la regresión (2.62). En [49] se describe el siguiente procedimiento de prueba:

1. Estimar el modelo (2.1). Obtener los residuales cuadrados $\hat{u}^2$ de cada observación.
2. Obtener la R cuadrada de la regresión (2.62).
3. Formar el estadístico F descrito en (2.63) y calcular el valor-p usando la distribución $F_{k,n-k-1}$. Si el valor-p es menor que el nivel de significancia elegido, se rechaza la hipótesis nula.

- ***Prueba de White***

  A la ecuación (2.62) se le agregan los cuadrados y los productos cruzados de todas las variables independientes con sus respectivos parámetros. Pero el hecho de aumentar variables al modelos hace que crezcan el número de parámetros y esto es una desventaja ya que usa muchos grados de libertad para un modelo que originalmente tiene pocas variables independientes.
  Tomando en cuenta que los valores ajustados están definidos por

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik}. \tag{2.64}$$

  Si elevamos estos valores al cuadrado tendremos todos los cuadrados y productos cruzados de las variables independientes. Con esta consideración se prueba la heterocedasticidad usando la ecuación

$$\hat{u}^2 = \delta_0 + \delta_1 \hat{y} + \delta_2 \hat{y}^2 + \epsilon. \tag{2.65}$$

  Por lo tanto se establece la hipótesis nula $H_0 : \delta_1 = \delta_2 = 0$, y se puede usar el estadístico F para demostrarla. Los pasos realizar esta prueba de *White* serían:

1. Estimar el modelo (2.1). Obtener los residuales $\hat{u}$, cuadrados $\hat{u}^2$ y los valores ajustados $\hat{y}$ y sus cuadrados $\hat{y}^2$.
2. Realizar la regresión de (2.65). Obtener la R cuadrada de la regresión $R^2_{\hat{u}^2}$.
3. Determinar el estadístico F, el valor-p usando la distribución $F_{2,n-3}$ y realizar el análisis correspondiente.

Normalidad

Con el supuesto 6, de normalidad, se determinaron las distribuciones de los estimadores $\hat{\beta}_i$, $1 \leq i \leq k$, y se construyeron las pruebas de hipótesis e intervalos de confianza. Para determinar si este supuesto se cumple existen varias pruebas como se describe en [9]. En este trabajo se presentarán tres pruebas:

- ***Prueba de Jarque-Bera (JB)***
  La prueba de Jarque-Bera es prácticamente una prueba de hipótesis conjunta de los parámetros S y K, que son los coeficientes de *asimetría* y de *curtosis* de los residuos de MCO, respectivamente. El estadístico de prueba se define por

$$JB = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right], \tag{2.66}$$

  donde n es el tamaño de la muestra. De acuerdo a Jarque-Bera el estadístico JB sintóticamente sigue una distribución χ^2 con dos grados de libertad [22].

- ***Prueba de Anderson-Darling***
  Según Anderson y Darling [1] para probar que los residuales $\hat{u}_j$ provienen de una distribución normal propone asumir que los residuales provienen de una distribución normal, ordenarlos de menor a mayor: $\hat{u}_1 \leq \hat{u}_2 \leq \cdots \leq \hat{u}_n$, donde n es el tamaño de la muestra. Sea $F(w)$ la función de distribución acumulada se define el estadístico A^2 como

$$A^2 = -n - \frac{1}{n}\sum_{j=1}^{n}(2j-1)\left[\ln(F(\hat{u}_j) + \ln(1 - F(\hat{u}_{n-j+1})\right]. \tag{2.67}$$

  Si el valor del estadístico es *demasiado* grande, se rechaza la hipótesis nula de normalidad. D'Agostino [8] calcula un estadístico modificado

$$A = A^2\left(1{,}0 + \frac{0{,}75}{n} + \frac{2{,}25}{n^2}\right), \tag{2.68}$$

y presenta los respectivos valores críticos; por lo tanto, se rechaza la hipótesis nula de normalidad si el estadístico A excede el valor crítico. En la Tabla (2.1) se presenta ciertos valores críticos para los principales niveles de confianza.

Cuadro 2.1: Niveles y sus valores críticos para la prueba de Anderson-Darling

Nivel de Confianza	Valor crítico
0,10	0,631
0,05	0,752
0,01	1,035

- ***Test de Shapiro-Wilk***

  El test de Shapiro-Wilk [43] establece como hipótesis nula que la muestra $y_1, \ldots, y_n$ proviene de una población con distribución normal. El estadístico W para la normalidad se define por

$$W = \frac{\left(\sum_{i=1}^{n} a_i y_1\right)^2}{\sum_{i=1}^{n} \left(y_i - \overline{y}\right)^2}, \tag{2.69}$$

  donde $a' = (a_1, \ldots, a_n) = \dfrac{m'V^{-1}}{(m'V^{-1}V^{-1}m)^{1/2}}$, siendo $m' = (m_1, \ldots, m_n)$ el vector de los valores esperados del estadístico de orden de una normal estándar y V la correspondiente matriz de covarianzas de orden n.

  Se rechaza la hipótesis nula, de normalidad, si el valor-p es menor que el nivel de significancia elegido, caso contrario no se puede rechazar dicha hipótesis.

2.1.4. Medidas de bondad de ajuste

Dado un modelo de regresión, nos interesa evaluar el rendimiento de este modelo, es decir, poder medir qué tan bien las variables independientes explican a la variable dependiente. En el presente trabajo usaremos las siguientes estadísticos:

- *El coeficiente de determinación*, denotado R^2, que se define como

$$R^2 = \frac{SCE}{SCT}, \tag{2.70}$$

  se interpreta como la proporción de la variación muestral de la variable dependiente y que es explicada por las variables independientes.

Una forma equivalente a (2.70), demostrada en [14] es

$$r^2 = \frac{\left[\sum_{i=1}^{n}\left(y_i - \bar{y}\right)\left(\hat{y}_i - \hat{\bar{y}}\right)\right]^2}{\left[\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2\right]\left[\sum_{i=1}^{n}\left(\hat{y}_i - \hat{\bar{y}}\right)^2\right]}, \tag{2.71}$$

es decir, es igual al cuadrado del coeficiente de correlación muestral entre y_i y $\hat{y}_i$. Para regresiones a través del origen tenemos la ecuación (2.72)[4].

$$R^2 = \frac{\sum_{i=1}^{n}\hat{y}_i^2}{\sum_{i=1}^{n}y_i^2}. \tag{2.72}$$

En la regresión múltiple $R = \sqrt{R^2}$ es llamado el *coeficiente de correlación múltiple* porque mide la relación entre la variable independiente y y el conjunto de variables independientes. El estadístico R^2 es usado para medir el ajuste y comparar entre modelos de regresión lineal múltiple y redes neuronales artificiales [2], [23], [32], [40].

- La *raíz del error medio cuadrático* ($RMSE$) es otro estadístico ampliamente usado para evaluar la bondad de ajuste, como una función de entrenamiento, con puntuación perfecta de 0. Aunque según cita Benjamin Schroeter [40] este estadístico puede ser ambiguo y no confiable ya que es una función de tres métricas separadas [23], [32]. Se define como

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}. \tag{2.73}$$

- El *error medio absoluto* (MAE) identifica mejor la magnitud del error y proporciona un panorama mejor del rendimiento del modelo, se esperaría como una puntuación perfecta el valor de 0. Una de las desventajas de este estadístico es que no discrimina entre sobre o subestimación [40]. Se defime como sigue:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2.74}$$

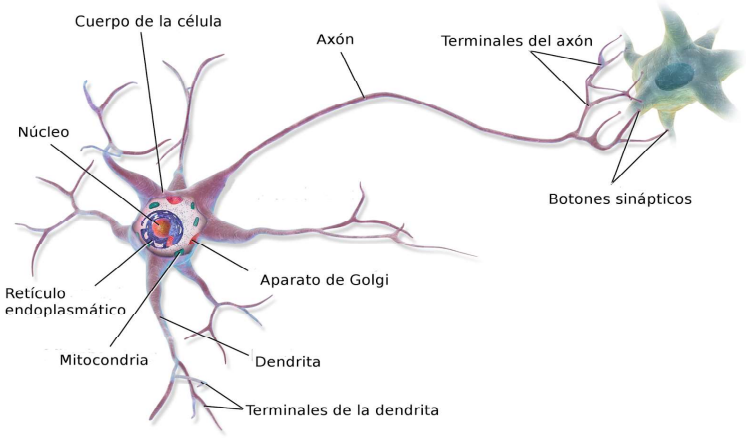
2.2. Redes neuronales artificiales

2.2.1. Neurona biológica

La célula fundamental del sistema nervioso es la neurona, y tienen como función la de conducir los estímulos eléctricos producto de reacciones físico químicas bajo ciertas condiciones. La neurona se puede dividir en tres partes principales: dentrita, cuerpo de la célula (soma) y el axón. La estructura de una neurona se presenta en la Figura 2.1 [1].

Las dendritas están compuestas por varias extensiones que forman el árbol dendrítico, donde se encuentran las terminales de la dendrina. Su principal función es adquirir continuamente los estímulos de otras neuronas o del medio exterior. El cuerpo de la célula es el responsable de procesar toda la información que proviene de las dendritas, y producir un potencial de activación que indica si la neurona puede desencadenar un impulso eléctrico a lo largo de su axón. También es el lugar donde se encuentran los órganos citoplasmáticos de la neurona, como el núcleo, la mitocondria, entre otros. El axón se compone de una única extensión cuya misión es la de guir los impulsos eléctricos a otras neuronas conectivas. La terminación del axón se compone de ramas llamadas terminales sinápticas.

Figura 2.1: Estructura de una neurona biológica



Las sinapsis son las conexiones que permiten la transferencia de los impulsos eléctricos de los axones de una neurona a las dendritas de otras neuronas. Hay que tener en cuenta que no hay contacto físico entre las neuronas que forman la unión sináptica, por lo tanto los elementos neurotransmisores liberados en la unión son los que están a cargo de ponderar la transmisión de una neurona a otra. En resumen, el funcionamiento de una neurona de

[1] **Fuente:** http://taycalbacete.blogspot.com/2014/05/la-computacion-neuronal.html

un cerebro humano es el siguiente: cuando la célula nerviosa se estimula con un impulso superior a su ***umbral de activación*** (-55 mV), se dispara un potencial eléctrico que se propagará a lo largo de su axón con una amplitud máxima de 35 mV (***potencial de acción***). La duración de esta señal es variable y depende de la categoría de la neurona, y una vez finalice el proceso de excitación el voltaje de acción volverá a su voltaje de reposo (-70 mV) [34].

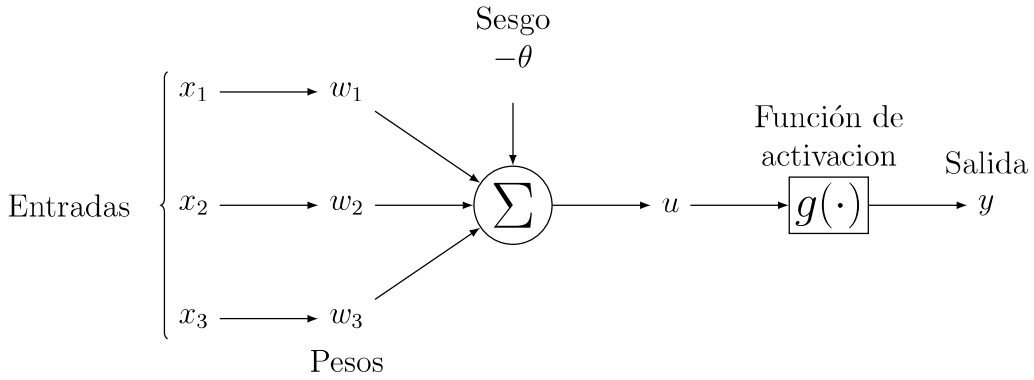
2.2.2. Neurona artificial

Una neurona artificial es un modelo simplificado de una neurona biológica y está inspirado en cómo una membrana celular de una neurona genera y propaga los impulsos eléctricos; las estructuras de redes neuronales artificiales (ANN) se desarrollaron a partir de los modelos conocidos de sistemas nerviosos biológicos y del propio cerebro humano.

Las neuronas artificiales usadas en las redes neuronales artificiales no son lineales y generalmente proporcionan salidas continuas, y realizan funciones simples como la de recopilación de señales disponibles en sus entradas y las ensamblan de acuerdo a sus funciones operativas y producen una respuesta teniendo en cuenta sus funciones de activación innatas.

El modelo propuesto por McCulloch y Pitts [34] es el más simple que incluye las características principales de una red neuronal biológica y es el modelo más utilizado en diferentes arquitecturas de redes neuronales artificiales.

Figura 2.2: Estructura de una neurona artificial



En la Figura 2.2 se muestra la estructura simplificada de una neurona artificial de tres señales de entrada y una señal de salida. Se considera que cada neurona está compuesta por siete elementos básicos [34]:

(a) Señales de entrada $(x_1, x_2, \ldots, x_n)$ son señales o muestras que provienen del entorno exterior y representan los valores que toman las variables de un experimento en particular. Generalmente las señales son normalizadas para mejorar la eficiencia computacional de los algoritmos de aprendizaje.

(b) Pesos sinápticos $(w_1, w_2, \ldots, w_n)$ son los valores utilizados para ponderar cada una de las variables de entrada, lo que permite cuantificar la relevancia de cada variable con respecto a la funcionalidad de la neurona.

(c) El agregador lineal (Σ) que reúne las señales de entrada ponderadas por los pesos sinápticos para producir un voltaje de activación.

(d) El umbral de activación o sesgo (θ) es una variable que se utiliza para determinar el umbral adecuado que el resultado producido por el agregador lineal debería tener para generar un valor de activación.

(e) El potencial de activación (u) que es el resultado de la diferencia entre el agregador lineal y el umbral de activación. Si este valor es positivo, es decir, si $u \geq 0$ entonces la neurona produce un potencial excitador,caso contrario será inhibitorio.

(f) Función de activación (g) cuyo objetivo es limitar la salida de la neurona dentro de un rango razonable de valores. La función puede ser parcialmente diferenciable (función de Heaviside, de paso bipolar o de rampa simétrica) o totalmente diferenciable (función logística, tangente hiperbólica, Gaussiana o lineal); su elección depende del tipo de red neuronal a usarse, como se detallará en las sección 2.2.4.

(g) La señal de salida (y) es el valor final producido por la neurona dado el conjunto de señales de entrada. También puede usarse como entrada de otras neuronas interconectadas secuencialmente.

Resumiendo el funcionamiento de la neurona, empezamos con la entrada de las señales representadas por las variables de entrada x_i con $i = 1, \ldots, n$; cada una de estas variables es multiplicada por su correspondiente peso sináptico obteniendo $w_i \cdot x_i$ con $i = 1, \ldots, n$. El potencial de activación u se obtiene restando de la suma de las variables ponderadas el valor del umbral de activación: $u = \sum_{i=1}^{n} w_i \cdot x_i - \theta$. Finalmente se aplica la función de activación al valor del potencial de activación, obteniendo $g(u)$.

En síntesis, el resultado producido por la neurona se resume en las ecuaciones (2,75) y (2,76).

$$u = \sum_{i=1}^{n} w_i \cdot x_i - \theta \tag{2.75}$$

$$y = g\,(u) \tag{2.76}$$

Las funciones de activación usadas en las redes neuronales artificiales se pueden dividir en dos grupos fundamentales: funciones parcialmente diferenciables y funciones totalmente diferenciables.

Las funciones de activación parcialmente diferenciables comúnmente usadas son:

(a) ***Función escalón o función de Heaviside***

El resultado obtenido con esta función dará valores unitarios positivos cuando el potencial de activación neuronal sea mayor o igual a cero, caso contrario, el resultado será nulo. Por lo tanto se tiene:

$$g(u) = \begin{cases} 1, \text{ si } u \geq 0 \\ 0, \text{ si } u < 0 \end{cases} \tag{2.77}$$

La representación de la función escalón se ilustra en la Figura 2.3 (a).

(b) ***Función de paso bipolar***

El resultado producido por esta función asumirá valores unitarios positivos cuando el potencial de activación de la neurona sea mayor que cero; un valor nulo cuando el potencial sea nulo y un valor unitario negativo cuando el potencial sea menor que cero:

$$g(u) = \begin{cases} 1, \text{ si } u > 0 \\ 0, \text{ si } u = 0 \\ -1, \text{ si } u < 0 \end{cases} \tag{2.78}$$

La representación de la función de paso bipolar se ilustra en la Figura 2.3 (b).

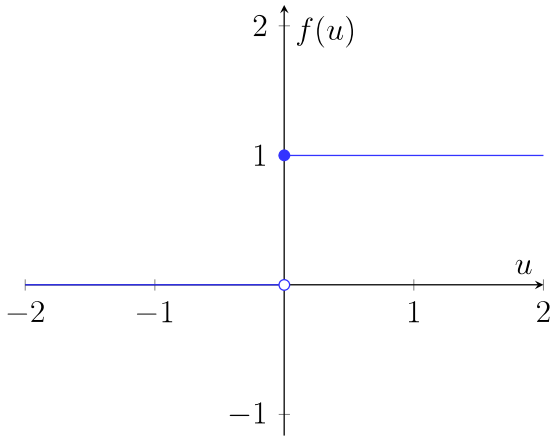
(c) ***Función de rampa simétrica***

Los valores que toma la función son iguales a los valores del potencial de activación cuando se definen dentro del rango $[-a, a]$,caso contrario toman los valores de sus límites:

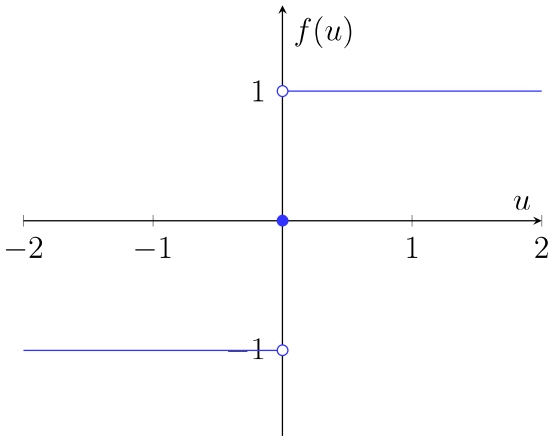
$$g(u) = \begin{cases} a, \text{ si } u > a \\ u, \text{ si } -a \leq u \leq a \\ -a, \text{ si } u < a \end{cases} \tag{2.79}$$

La representación de la función de rampa simétrica se ilustra en la Figura 2.3 (c).

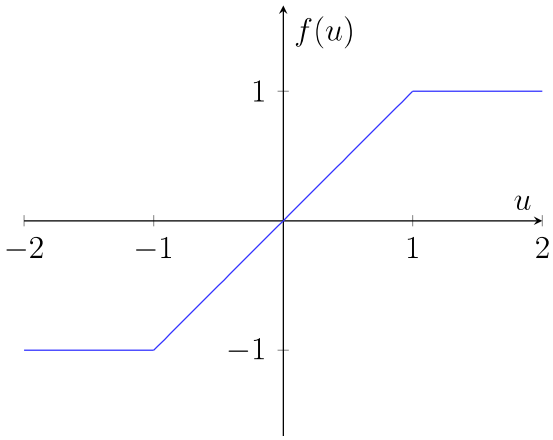
Figura 2.3: Funciones de activación parcialmente diferenciables



(a) Función escalón o función de Heaviside



(b) Función de paso bipolar



(c) Función de rampa simétrica, con $a = 1$

Las funciones de activación totalmente diferenciables más usadas son:

(a) ***Función logística***

El resultado producido por la función logística siempre estará entre cero y uno, expresada matemáticamente por:

$$g(u) = \frac{1}{1 + e^{-\beta \cdot u}}, \tag{2.80}$$

donde β es una constante asociada con la pendiente de la función en su punto de inflexión. La representación de la función logística se ilustra en la Figura 2.4 (a). En la Figura 2.4 (b) se muestra cómo varía la función logística cuando varía el parámetro β, la función toma un tono más azul mientras más grande es el valor del parámetro.

(b) ***Función tangente hiperbólica***

Los valores que toma esta función están siempre entre -1 y 1, se expresa matemáticamente de la siguiente manera:

$$g(u) = \frac{1 - e^{-\beta \cdot u}}{1 + e^{-\beta \cdot u}}, \tag{2.81}$$

donde el parámetro β está asociado con la pendiente de la función en su punto de inflexión. La representación de la función tangente hiperbólica se ilustra en la Figura 2.4 (c). En la Figura 2.4 (d) se muestra cómo varía la función cuando varía el parámetro β, la función toma un tono más azul mientras más grande es el valor del parámetro.

(c) ***Función Gaussiana***

En el caso de funciones de activación Gaussianas la salida de las neuronas producirá resultados iguales para valores de potencial de activación u ubicados a la misma distancia de su valor promedio. Está expresada de la siguiente manera:

$$g(u) = e^{-\frac{(u-c)^2}{2\sigma^2}}, \tag{2.82}$$

donde c es el centro de la función Gaussiana y σ es su desviación estándar. La representación gráfica de esta función se encuentra en la Figura 2.4 (e).

(d) ***Función lineal***

La función de activación lineal produce resultados iguales a su potencial de activación u, y tiene la siguiente expresión:

$$g(u) = u \tag{2.83}$$

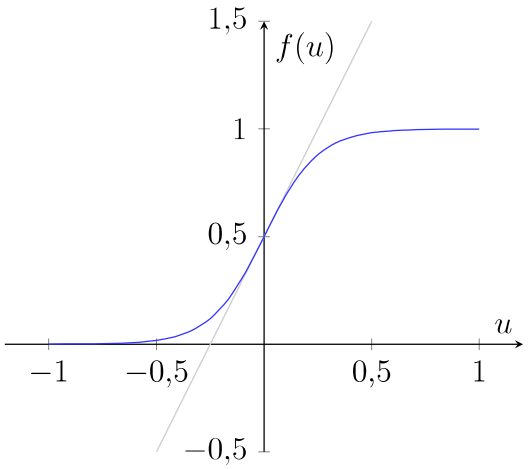
La representación gráfica de esta función se encuentra en la Figura 2.4 (f).

2.2.3. Arquitecturas de Redes Neuronales Artificiales y Procesos de Aprendizaje

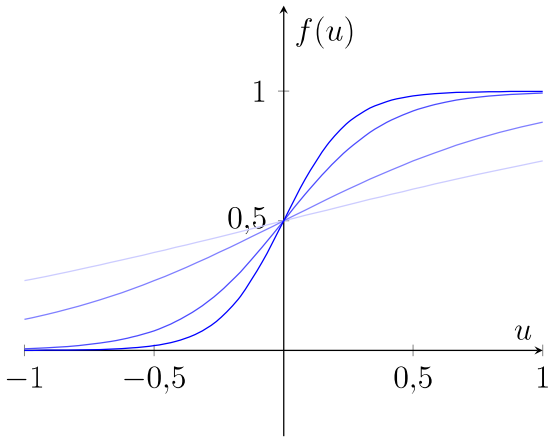
Arquitecturas

Las arquitecturas de las redes neuronales artificiales son las que definen cómo las neuronas están dispuestas o colocadas en relación a cada una de ellas, esta estructura es la que dirige las conexiones sinápticas de las neuronas. Por otro lado, el proceso de entrenamiento implica aplicar un conjunto de pasos ordenados (algoritmos de aprendizaje) para

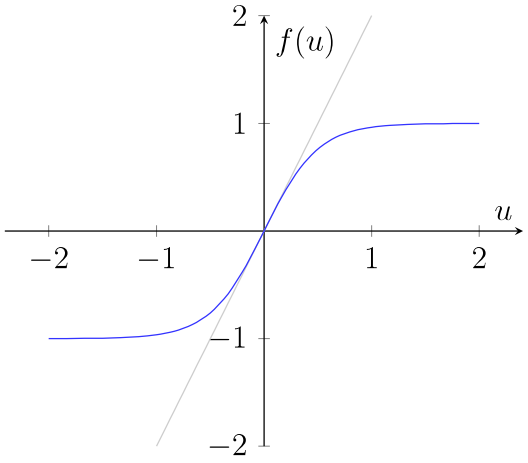
Figura 2.4: Funciones de activación totalmente diferenciables



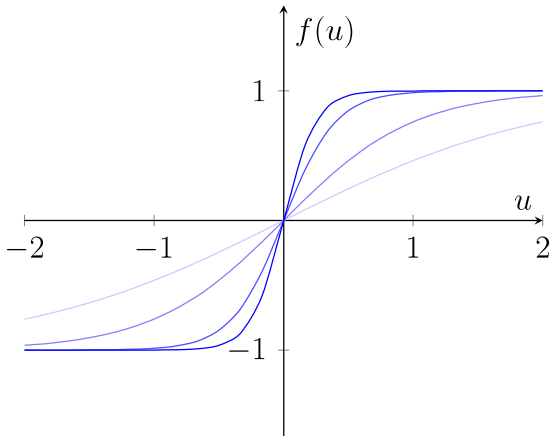
(a) La función de activación logística



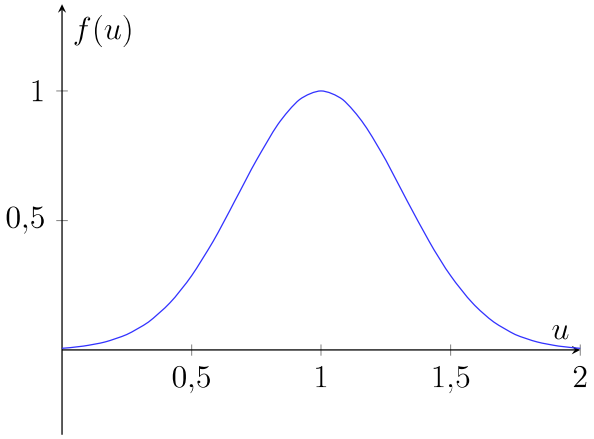
(b) La función de activación logística variando β



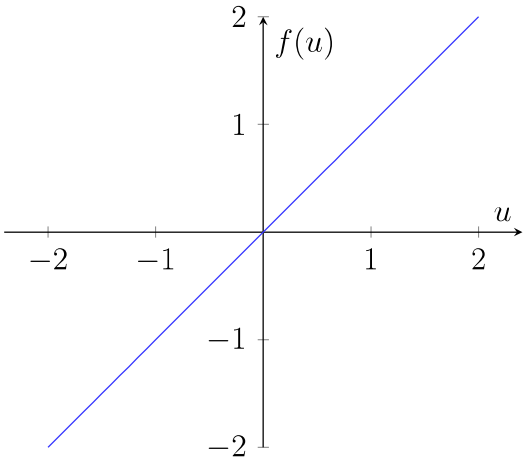
(c) La función de activación tangente hiperbólica



(d) La función de activación tangente hiperbólica variando β



(e) La función de activación Gaussiana



(f) La función de activación lineal

ajustar los pesos y los umbrales de las neuronas para que sus salidas estén cerca de los valores deseados.

En general una red neuronal artificial está dividida en tres partes llamadas capas, capa de entrada, capas ocultas y capa de salida:

(a) ***Capa de entrada***
Esta capa es la responsable de recibir información (datos, muestras o patrones), que generalmente se normaliza dentro de valores límite producidos por las funciones de activación. Esta normalización da como resultado una mejor precisión numérica para las operaciones matemáticas realizadas por la red.

(b) ***Capas ocultas***
Estas capas están compuestas por neuronas que son responsables de extraer patrones asociados con el proceso o sistema que se está analizando. Estas capas realizan la mayor parte del procesamiento interno de una red.

(c) ***Capas de salida***
Esta capa es la responsable de producir y presentar las salidas de la red, que son el resultado del proceso realizado en las capas anteriores.

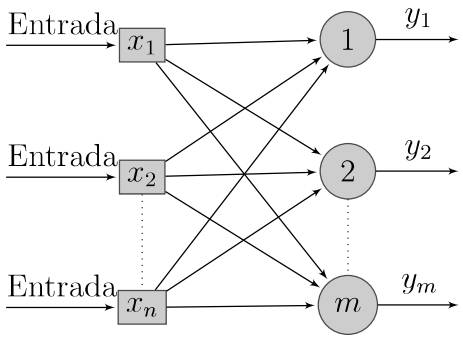
Las principales arquitecturas de redes neuronales artificiales se pueden dividir en: (1) red prealimentada de una capa, (2) red prealimentada de múltiples capas, (3) red recurrente y (4) red de malla.

1. ***Arquitectura prealimentada de una capa***
Esta red neuronal artificial tiene una sola capa de entrada y una sola capa neuronal que es también la capa de salida. En la Figura 2.5 se presenta una red prealimentada de n entradas y m salidas; la información fluye en un solo sentido que es desde la capa de entrada a la capa de salida.

   Las principales redes de este tipo de arquitectura son la red Perceptrón (Figura 2.7) y la red ADALINE (Figura 2.8) cuyos algoritmos de aprendizaje usados en sus procesos de entrenamiento se basan en la regla de Hebb y la regla Delta respectivamente [34]. En la subsección 2.2.4 se detallará cada una de estas dos redes y el funcionamiento de sus respectivos algoritmos de entrenamiento.

2. ***Arquitectura prealimentada multicapa***
Este tipo de red neuronal posee una o más capas ocultas, un ejemplo de este tipo

Figura 2.5: Red prealimentada de una capa



de arquitectura se presenta en la Figura 2.6, con n señales de entrada y dos capas neuronales ocultas con n_1 y n_2 neuronas respectivamente y una capa de salida con m neuronas que representan los respectivos valores de salida del problema analizado. Entre las diversas aplicaciones está la aproximación de funciones, clasificación de patrones, optimización, robótica, etc [34].

Las principales redes prealimentadas multicapa son el Perceptrón Multicapa (MLP: Multilayer Perceptron) y la Función de Base Radial (RBF: Radial Basis Function) cuyos algoritmos de aprendizaje usados en sus procesos de entrenamiento son la regla delta generalizada y la regla delta, respectivamente. En la subsección 2.2.5 se entrará a fondo en el estudio de las redes Perceptrón Multicapa y sus algoritmos de entrenamiento.

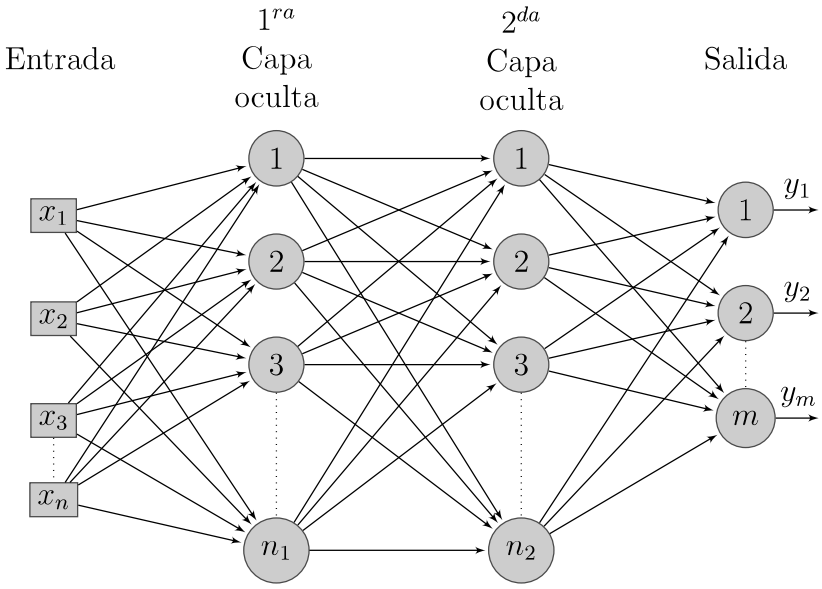
Existen otras estructuras de redes neuronales que no serán tratadas en el presente trabajo, tales como la *Arquitectura recurrente o de retroalimentación* y la *Arquitectura de Malla* [34].

Procesos de aprendizaje

Como se mencionó anteriormente el proceso de entrenamiento de una red neuronal consiste en aplicar los pasos ordenados requeridos para ajustar los pesos y umbrales sinápticos de sus neuronas, a fin de generalizar las soluciones producidas por sus salidas. Este conjunto de pasos ordenados utilizados para entrenar la red se llama algoritmo de aprendizaje. Entre los distintos tipos de aprendizaje tenemos:

1. ***Aprendizaje supervisado***
   Esta estrategia de aprendizaje parte de un conjunto en el cual se disponen de las variables de entrada o variables independientes y de sus correspondientes respuestas

Figura 2.6: Ejemplo de una red prealimentada multicapa



o variables dependientes. La estrategia consiste en determinar el comportamiento de los datos de entrada con el fin de obtener los datos de salida. Ciertos autores tratan a la variable respuesta como a un *profesor*, que muestra cuál es la respuesta correcta que la red debe obtener a partir de los datos de entrada [38].

Los pesos y umbrales sinápticos de la red se ajustan continuamente a través de la aplicación de acciones comparativas, ejecutadas por el propio algoritmo de aprendizaje, que supervisa la diferencia entre las salidas producidas con respecto a las salidas deseadas, utilizando esta diferencia en el procedimiento de ajuste. La red se considera *entrenada* cuando esta diferencia se encuentra dentro de un rango propuesto[34].

2. ***Aprendizaje no supervisado***

   El aprendizaje no supervisado no requiere de la variable respuesta, es decir, solo se disponen de los datos de entrada y la red aprende alguna característica interna o patrón del conjunto de datos de entrada para derivar una salida [5]. Un algoritmo basado en un aprendizaje no supervisado se organiza de acuerdo a particularidades de los elementos de la muestra, identificando subconjuntos que presentan similitudes. El algoritmo de aprendizaje ajusta los pesos y umbrales sinápticos de la red para reflejar estos grupos dentro de la propia red [34].

Ciaburro y Venkateswaran [5] describen las aplicaciones principales que se desarrollan con los dos tipos de aprendizaje:

- Aprendizaje supervisado: regresión y clasificación.

- Aprendizaje no supervisado: clusterización.

En adelante, se entenderá como **entrenamiento supervisado** a la aplicación de un algoritmo de aprendizaje supervisado, y **entrenamiento no supervisado** a la aplicación de un algoritmo de aprendizaje no supervisado.

Normalización

Priddy y Keller [38] recomiendan normalizar los datos de entrada para acelerar el proceso de entrenamiento, en especial cuando las entradas se encuentran en escalas diferentes; entre los tipos de normalización que suguieren tenemos: la normalización estadística o normalización $Z-$Score, la normalización *Min-Max*, sigmoidal o normalización softmax, normalización energía y la normalización por componentes principales. Para una mejor comparación entre el modelo de redes nuronales artificiales y el de regresión lineal múltiple Niazian [32] recomienda ejecutar ambos modelos con los mismos datos normalizados.

El nuevo rango de valores para la normalización depende de la función de activación a usarse en la red, bajo este principio Shanmuganathan & Samarasinghe [41], Nunes da Silva, *et al* [34], Priddy & Keller [38] y Ciaburro & Venkateswaran [5] establecen los intervalos $[0, 1]$ y $[-1, 1]$ como los nuevos rangos para las funciones de activación logística y tangente hiperbólica, respectivamente.

La normalización *Min-Max* es la comúnmente usada en los modelos de redes neuronales, ya que se puede normalizar al rango de datos que se desee, que normalmente son los rangos de las funciones de activación, $[0, 1]$ o $[-1, 1]$. La fórmula de normalización es la siguiente:

$$x'_i = (max_t - min_t) \times \left(\frac{(x_i - min_v)}{(max_v - min_v)} \right) + min_t, \tag{2.84}$$

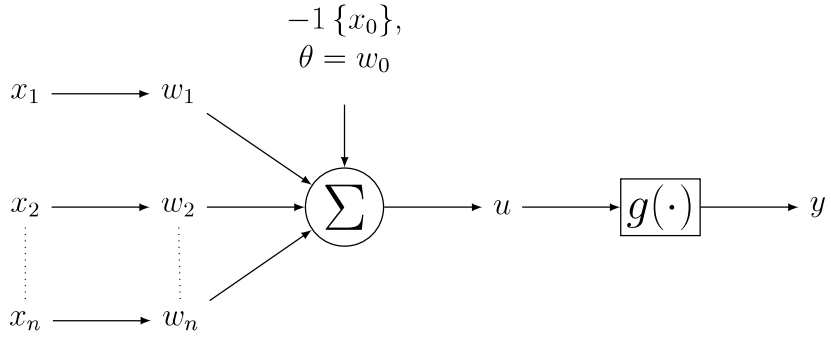
donde x'_i es el valor normalizado, x_i el valor real, min_v y max_v son los valores mínimo y máximo del conjunto de datos de entrada; y, min_t y max_t son los valores mínimo y máximo del nuevo rango de los datos. Hay que considerar que $max_v - min_v \neq 0$ y que si $max_v - min_v = 0$ entonces se tiene que la entrada tiene un valor constante, y debería quitarse pues no aporta ninguna información a la red neuronal [38].

2.2.4. Redes prealimentadas de una capa

Red Perceptrón

La red Perceptrón pertenece a la clase de arquitecturas prealimentadas de una sola capa, ya que la información fluye en su estructura desde la capa de entrada hasta la capa de salida que posee una única neurona. Generalmente la red Perceptrón es usada para reconocimiento de patrones o para problemas de clasificación. En la figura 2.7 se presentan los componentes de la red Perceptrón.

Figura 2.7: La red Perceptrón



Como se mencionó en la sección 2.1.2 la red Perceptrón se puede describir con las ecuaciones 2.75 y 2.76, y las funciones de activación usualmente usadas son la función escalón y la función de paso bipolar presentadas en la secci ón 2.2.2. Por lo tanto, indiferentemente de la función de activación usada, la salida de la red tomará dos valores: 0 y 1 para la función escalón tal como se define en la ecuación (2.77); y -1 y 1 para la función bipolar como en la ecuación (2.78).

Sin pérdida de generalidad, en un problema de clasificación, la red Perceptrón clasifica las señales de entrada en dos posibles clases, la clase A y la clase B. Independientemente, podemos asignar que cuando la salida sea -1 signifique pertenecer a la clase A, y que cuando la salida sea 1 signifique pertenecer a la clase B, esto cuando la función de activación es la bipolar.

Si bien los valores de las muestras o señales de entrada pueden asumir cualquier valor numérico, en la práctica se usan técnicas de normalización para mejorar el rendimiento computacional y depende del dominio de la función de activación elegida.

El ajuste de los pesos y el umbral en el Perceptrón se realiza a través de un entrenamiento supervisado, descrito en la subsección 2.2.3. Dado que el Perceptrón se usa

generalmente en problemas de reconocimiento de patrones, y teniendo en cuenta que su salida puede asumir solo dos valores posibles, entonces cada salida se asocia con una de las dos clases que están identificandas y que deben ser linealmente separables; estas condiciones son las establecidas en el *Teorema de convergencia del Perceptrón* descrita por Minski M. y Papert S., la demostración de este teorema queda fuera del alcance del presente trabajo, pero se encuentra en [26], página 167.

El ajuste de los pesos y umbrales de la red Perceptrón se realiza mediante la regla de aprendizaje de Hebb. Hay que tener en cuenta que Donald Hebb enuncia un postulado neurofisiológico, acerca de cómo se conectan las neuronas biológicas y que la regla de aprendizaje de Hebb es una adaptación matemática de este postulado [16], [15]. Matemáticamente, sean x_i el valor de entrada de la neurona pre-sináptica y y_i el valor de salida de la neurona pre-sináptica, entonces la fuerza de la sinapsis que conecta la neurona j y la neurona i, representada por w_{ij}, se define como

$$w_{ij} = \frac{1}{N} x_i y_i, \tag{2.85}$$

donde N es el número de neuronas en la red. Tomando incrementos continuos de la regla de Hebb se tiene

$$w_{ij}^{nuevo} = w_{ij}^{anterior} + \eta x_i y_i, \tag{2.86}$$

donde η es una constante positiva llamada tasa de aprendizaje [17], [15].

Antes de presentar el Algoritmo de Perceptrón consideramos las siguientes notaciones: el vector de los pesos y el umbral $\boldsymbol{w} = [\theta, w_1, w_2, \cdots, w_n]^T$, la k-ésima muestra para el entrenamiento $\boldsymbol{x}^{(k)} = \left[-1, x_1^{(k)}, x_2^{(k)}, \cdots, x_n^{(k)}\right]^T$, la salida del Perceptrón y, $d^{(k)}$ es el valor de salida para la k-ésima muestra para el entrenamiento, y η que define la tasa de aprendizaje del Perceptrón y generalmente está en el rango $0 < \eta < 1$. El proceso de entrenamiento de la red Perceptrón se describe en el Algoritmo 1.

La red se considera entrenada cuando no se encuentran errores entre los valores reales y los producidos por las salidas de la red. Luego de que la red está entrenada aplicamos el Algoritmo 2 para clasificar nuevas muestras.

En los casos prácticos hay que tener en cuenta que el algoritmo diverge cuando el problema no es linealmente separable, es decir, no existe un hiperplano que separe las dos clases.

Algoritmo 1: Algoritmo Perceptrón: Entrenamiento

Entrada: Un conjunto de entrenamiento $\left\{\mathbf{x}^{(k)}\right\}$

Salida: El vector $\mathbf{w}$ de los pesos y el umbral que ajustan la red Perceptrón

Asociar cada salida deseada $\left\{d^{(k)}\right\}$ a cada muestra
Inicializar $\mathbf{w}$ con valores aleatorios pequeños
Definir la tasa de aprendizaje η
Inicializar un contador de iteraciones $c \leftarrow 0$
repetir
    $\varepsilon \leftarrow 0$
    $\epsilon \leftarrow 0$
    para *todo* k **hacer**
        $u \leftarrow \mathbf{w}^T \cdot \mathbf{x}^{(k)}$
        $y \leftarrow f(u)$
        si $y \neq d^{(k)}$ **entonces**
            $\mathbf{w} \leftarrow \mathbf{w} + \eta(d^{(k)} - y) \cdot \mathbf{x}^{(k)}$
            $\epsilon = 1$, *aún existe error en las salidas.*
    $c \leftarrow c + 1$
hasta que $\varepsilon = \epsilon$
devolver $\mathbf{w}$

Algoritmo 2: Algoritmo Perceptrón: Operación

Entrada: Un conjunto a ser clasificado $\{\mathbf{x}\}$ que puede pertenecer a la clase A o a la clase B

Usar el vector $\mathbf{w}$ ajustado con el Algoritmo 1
para $\{\mathbf{x}\}$ **hacer**
    $u \leftarrow \mathbf{w}^T \cdot \mathbf{x}$
    $y \leftarrow f(u)$
    si $y = -1$ **entonces**
        $\mathbf{x} \in A$
    si $y = 1$ **entonces**
        $\mathbf{x} \in B$

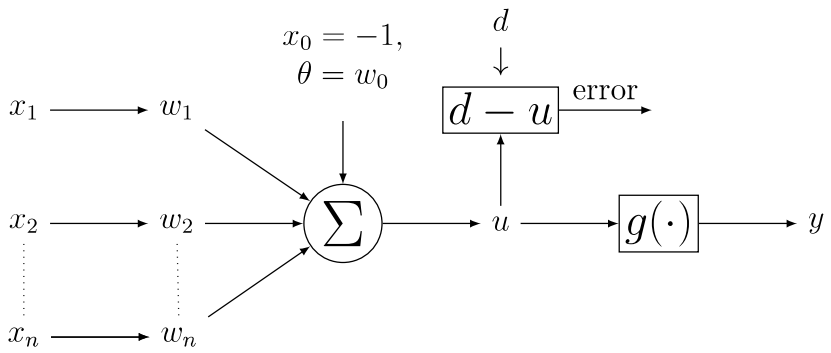
Red ADALINE

La red ADALINE (Adaptive Linear Element) también pertenece a la clase de arquitecturas prealimentadas de una sola capa, porque la información fluye desde la capa de entrada hacia la capa de salida. Al igual que la red Perceptrón, la red ADALINE es usada principalmente el problemas de clasificación de patrones que involucran solo dos clases, la clase A y la clase B, que son asociadas con las posibles salidas que produce la función de activación que puede ser la función de escalón o la función de paso bipolar.

Se puede observar en la figura 2.8 que la estructura de la red ADALINE es similar a la del Perceptrón, a diferencia de un bloque en la estructura ADALINE cuya función es ayudar con el proceso de entrenamiento de la red. Este bloque produce la señal de error que se define como

$$error = d - u. \tag{2.87}$$

Figura 2.8: Red ADALINE



El proceso de entrenamiento de la red ADALINE se basa en el algoritmo de aprendizaje denominado Regla Delta, que consiste en minimizar la diferencia entre la salida d y el potencial de activación u, considerando p muestras para ajustar los pesos y el umbral de la neurona.

***Definición* 2.6.** *El error cuadrático de p muestras de entrenamiento se define por*

$$E(\mathbf{w}) = \frac{1}{2}\sum_{k=1}^{p}\left(d^{(k)} - u\right)^2, \tag{2.88}$$

donde $d^{(k)}$ es la salida esperada de la k-ésima muestra y u es tal como se define en la ecuación (2.75).

***Definición* 2.7.** *El error cuadrático medio con respecto a p muestras de entrenamiento*

se define por

$$\bar{E}\left(\mathbf{w}\right) = \frac{1}{p}\sum_{k=1}^{p}\left(d^{(k)} - u\right)^2. \tag{2.89}$$

***Definición* 2.8.** *El vector de pesos óptimo* $\mathbf{w}^*$ *se define como*

$$E\left(\mathbf{w}^*\right) \leq E\left(\mathbf{w}\right), \quad \forall \mathbf{w} \in \mathbb{R}^{n+1}. \tag{2.90}$$

Sustituyendo la ecuación (2.75) en la ecuación (2.88) tenemos

$$\begin{aligned} E(\mathbf{w}) &= \frac{1}{2}\sum_{k=1}^{p}\left(d^{(k)} - \left(\sum_{i=1}^{n} w_i \cdot x_i^{(k)} - \theta\right)\right)^2 \\ &= \frac{1}{2}\sum_{k=1}^{p}\left(d^{(k)} - \left(\mathbf{w}^T \cdot \mathbf{x}^{(k)} - \theta\right)\right)^2 \end{aligned} \tag{2.91}$$

Aplicamos el operador gradiente al error medio cuadrático con respecto al vector $\mathbf{w}$ para buscar el valor óptimo de la función de error cuadrático:

$$\begin{aligned} \nabla E\left(\mathbf{w}\right) &= \frac{\partial E\left(\mathbf{w}\right)}{\partial \mathbf{w}} \\ &= \sum_{k=1}^{p}\left(d^{(k)} - \left(\mathbf{w}^T \cdot \mathbf{x}^{(k)} - \theta\right)\right) \cdot \left(-\mathbf{x}^{(k)}\right) \\ &= -\sum_{k=1}^{p}\left(d^{(k)} - u\right) \cdot \mathbf{x}^{(k)} \end{aligned}$$

Dado que el objetivo de la optimización es minimizar el error cuadrático, se debe considerar el opuesto del vector gradiente en la actualización del vector de pesos, es así que la variación $\Delta\mathbf{w}$ está dada por

$$\begin{aligned} \Delta\mathbf{w} &= -\eta\nabla E\left(\mathbf{w}\right) \\ &= \eta\sum_{k=1}^{p}\left(d^{(k)} - u\right) \cdot \mathbf{x}^{(k)}, \end{aligned} \tag{2.92}$$

donde η es la tasa de aprendizaje, y define qué tan rápido avanza el proceso de entrenamiento de la red en dirección del punto que minimiza la función de error cuadrático. Generalmente se tiene que $0 < \eta < 1$.

Expresamos la ecuación (2.92) de manera que obtenemos el vector $\mathbf{w}$ actualizado

$$\mathbf{w}^{actual} = \mathbf{w}^{previo} + \eta \sum_{k=1}^{p} \left(d^{(k)} - u\right) \cdot \mathbf{x}^{(k)}, \tag{2.93}$$

considerando la actualización del vector $\mathbf{w}$ para cada k-ésima muestra de entrenamiento se tiene

$$\mathbf{w}^{actual} = \mathbf{w}^{previo} + \eta \left(d^{(k)} - u\right) \cdot \mathbf{x}^{(k)}, \quad k = 1, \ldots, p. \tag{2.94}$$

El algoritmo converge cuando la diferencia del error medio cuadrático entre dos actualizaciones sucesivas es lo suficientemente pequeña, es decir:

$$\left|\bar{E}\left(\mathbf{w}^{actual}\right) - \bar{E}\left(\mathbf{w}^{previo}\right)\right| \leq \varepsilon, \tag{2.95}$$

donde ε es la precisión de convergencia.

Presentamos el algoritmo de entrenamiento de la red ADALINE:

Algoritmo 3: Algoritmo ADALINE: Entrenamiento

Entrada: Un conjunto de entrenamiento $\left\{\mathbf{x}^{(k)}\right\}$

Salida: El vector $\mathbf{w}$ de los pesos y el umbral que ajustan la red ADALINE

```
Asociar cada salida {d^(k)} a cada muestra
Inicializar w con valores aleatorios pequeños
Especificar la tasa de aprendizaje η y la precisión ε
Inicializar un contador c ← 0
repetir
    Ē^previo ← Ē(w)
    para {x^(k), d^(k)} hacer
        u ← w^T · x^(k)
        w ← w + η · (d^(k) − u) · x^(k)
    c ← c + 1
    Ē^actual ← Ē(w)
hasta que |Ē^actual − Ē^previo| ≤ ε
devolver w
```

La variable c es un contador que registra el número de veces que es necesario tratar el conjunto de entrenamiento para ajustar el vector de pesos $\mathbf{w}$.

El algoritmo usado para calcular el error cuadrático medio se presenta en el siguiente

pseudocódigo:

Algoritmo 4: Algoritmo MSE

1 Obtener el número de muestras de entrenamiento p
2 Inicializar la variable $\bar{E} \leftarrow 0$
3 **para** $\left\{\mathbf{x}^{(k)}, d^{(k)}\right\}$ **hacer**
4 $\quad u \leftarrow \mathbf{w}^T \cdot \mathbf{x}^{(k)}$
5 $\quad \bar{E} \leftarrow \bar{E} + \left(d^{(k)} - u\right)^2$
6 $\bar{E} \leftarrow \dfrac{\bar{E}}{p}$

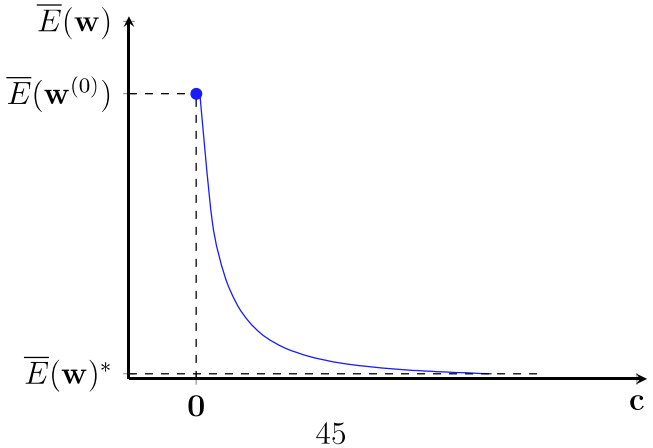
Una vez terminado el proceso de entrenamiento podemos clasificar nuevas muestras siguiendo los pasos que se dan a continuación:

Algoritmo 5: Algoritmo ADALINE: Operación

Entrada: Una muestra a ser clasificada $\{\mathbf{x}\}$
1 Usar el vector $\mathbf{w}$ ajustado con el Algoritmo 3
2 **para** $\{\mathbf{x}\}$ **hacer**
3 $\quad u \leftarrow \mathbf{w}^T \cdot \mathbf{x}$
4 $\quad y \leftarrow f(u)$
5 $\quad$ **si** $y = -1$ **entonces**
6 $\quad\quad \mathbf{x} \in A$
7 $\quad$ **si** $y = 1$ **entonces**
8 $\quad\quad \mathbf{x} \in B$

En la Figura 2.9 se muestra que la curva del error cuadrático medio del algoritmo ADALINE es siempre descendente, es decir, disminuye a medida que aumenta el número de iteraciones.

Figura 2.9: La función de error cuadrático medio con respecto al número de iteraciones



Comparación entre la red Perceptrón y la red ADALINE

Como se vio anteriormente, la red Perceptrón es ajustada mediante la regla de Hebb y la red ADALINE es ajustada mediante la regla Delta y ambas son usadas para clasificar las señales de entrada en dos posibles clases en la salida, la clase A y la clase B.

Pueden existir infinitos hiperplanos que separan la clase A y la clase B, cada uno de estos es válido pero existe un único hiperplano que es el óptimo y en esto radica la principal deferencia entre los procesos de entrenamiento Perceptrón y ADALINE: la solución hallada por la red Perceptrón dependen de los valores iniciales del vector de pesos por lo que para cada vector inicial la solución puede ser distinta, y no necesariamente es la solución óptima; por otro lado, independientemente de los valores iniciales del vector de pesos, la solución obtenida por la red ADALINE es la solución óptima y siempre será la misma.

Por lo expuesto, la red Perceptrón tiene mayor probabilidad de realizar una clasificación errónea cuando las entradas generan una salida que se encuentra en la frontera que separa la clase A con la clase B. Por otro lado, la red ADALINE reduce la probabilidad de clasificación incorrecta dado que usa el hiperplano óptimo obtenido mediante la minimización del error medio cuadrático [34].

Estos dos tipos de arquitecturas son la base para otras más complejas, que no solo admite una una capa intermedia que es la misma capa de salida y con una sola neurona que admite solo dos tipos de salida. Como se verá en la sección 2.2.5, las redes prealimentadas multicapa admiten varias capas ocultas y varias neuronas en la capa de salida que pueden tomar valores continuos en un cierto rango de salida.

2.2.5. Redes prealimentadas multicapa

La red perceptrón multicapa (MLP) presenta al menos una capa neuronal oculta, que se coloca entre la capa de entrada y la capa de salida, por lo que este tipo de red tiene al menos dos capas neuronales y sus neuronas se ubican entre las capas intermedias y la capa de salida.

Como se muestra en la Figura 2.6, el flujo de la red empieza por la capa de entrada, pasa a través de las capas intermedias y termina en la capa de salida. La información se transmitirá capa por capa, donde las salidas de las neuronas de la primera capa serán las entradas de las neuronas de la segunda capa, las neuronas de la segunda capa oculta serán

las entradas de las neuronas de la capa neuronal de salida. Por lo tanto, la propagación de las señales en una red MLP, independiente del número de capas, fluye en una sola dirección, desde la capa de entrada a la capa neuronal de salida.

A diferencia de las redes Perceptrón y ADALINE, la capa de salida de las redes MLP pueden estar compuestas de varias neuronas y cada una de estas neuronas representa una de las salidas del proceso que se está estudiando; por lo tanto, si el proceso tiene m salidas, entonces la red MLP tendrá m neuronas en su capa de salida. El algoritmo de aprendizaje utilizado durante el proceso de entrenamiento de un MLP se llama propagación hacia atrás (backpropagation).

Algoritmo Backpropagation (backprop)

Las etapas del algoritmo backpropagation se presentan en la Figura 2.10. La primera etapa se llama propagación hacia adelante, en esta etapa se inserta el conjunto de entrenamiento $\{x_1, x_2, \ldots, x_n\}$ que se propaga capa por capa hasta la capa de salida. En esta etapa unicamente se obtiene las respuestas de la red con los valores actuales de los pesos y umbrales que permanecen sin modificación durante esta etapa. Obtenidas las respuestas producidas por las neuronas se calculan las desviaciones (errores) respecto de las respuestas deseadas.

Debido a los errores se aplica la segunda parte del algoritmo backpropagation, conocida como propagación hacia atrás. En esta etapa se realizan los ajustes de los pesos y umbrales de todas las neuronas de la red MLP. La aplicación sucesiva de estas dos etapas permite que en cada iteración se ajusten los parámetros y disminuya la suma de los errores producidos entre las respuestas de la red con las respuestas deseadas.

La notación usada para desarrollar el algoritmo de propagación hacia atrás se ilustra en la Figura 2.11. Notaremos $W_{ji}^{(L)}$ como la matriz de pesos cuyos valores son los pesos que conectan a la neurona j de la capa L a la neurona i de la capa $(L-1)$. En la Figura 2.12 se ilustran los pesos de entrada para la neurona j de la capa L.

Figura 2.10: Ilustración de las dos etapas de una red MLP

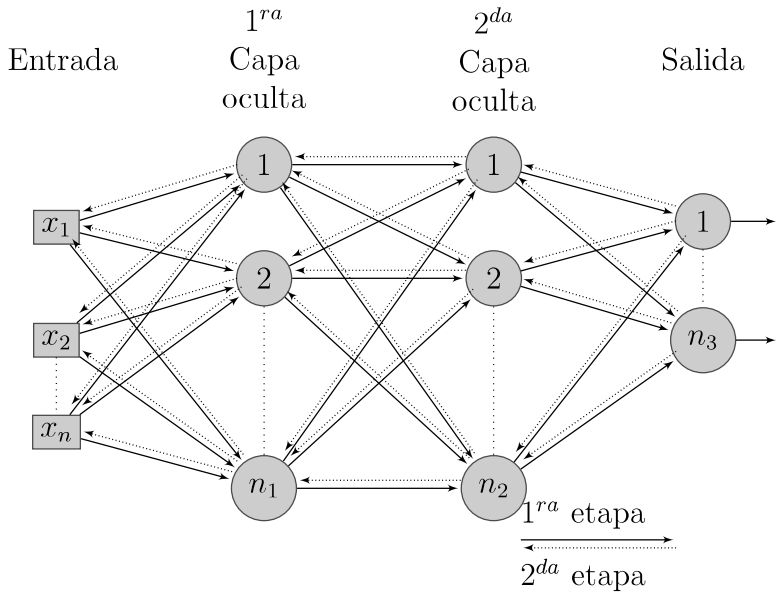


Figura 2.11: Notación usada en el algoritmo backpropagation

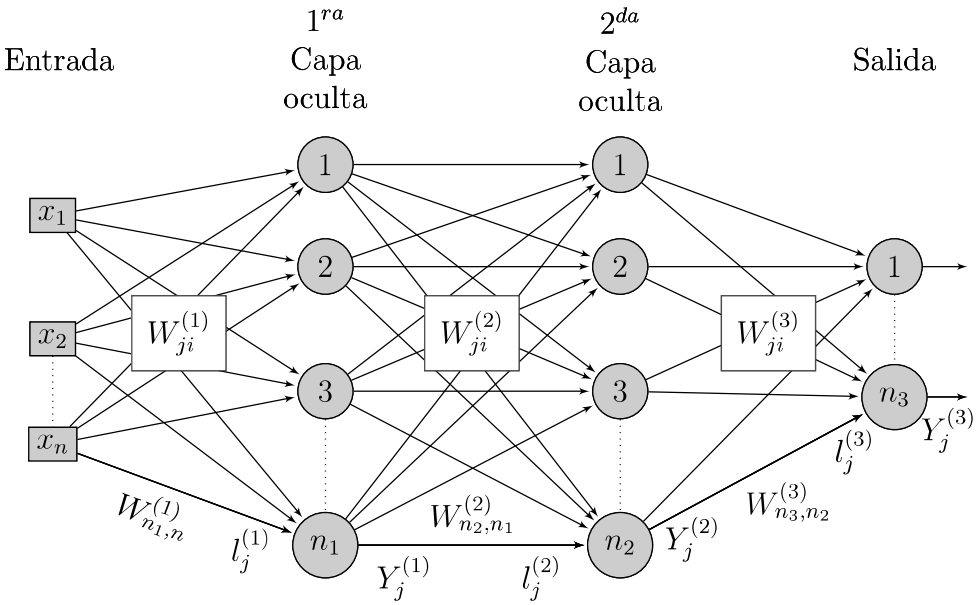
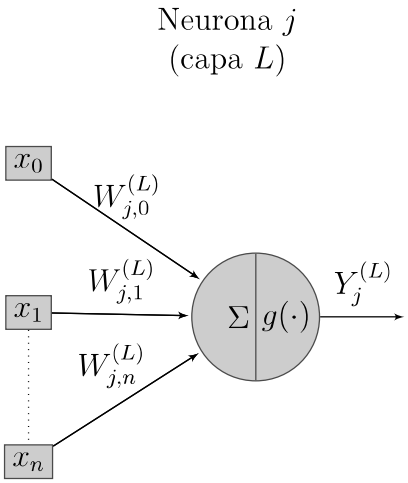


Figura 2.12: Ilustración de una neurona usada en el algoritmo backpropagation



***Definición* 2.9.** *El vector de entradas ponderadas $I_j^{(L)}$ relacionado a la neurona j de la capa L se define por*

$$I_j^{(L)} = \begin{cases} \sum_{i=0}^{n} W_{ji}^{(L)} \cdot x_i, \text{ si } L = 1 \\ \\ \sum_{i=0}^{n_L} W_{ji}^{(L)} \cdot Y_i^{(L-1)}, \text{ caso contrario,} \end{cases} \tag{2.96}$$

donde n_L es el número de nodos en la capa $(L-1)$.

***Definición* 2.10.** *El vector de salidas Y_j^L relacionado a la neurona j de la capa L se define por*

$$Y_j^L = g\left(I_j^L\right), \tag{2.97}$$

donde $g(\cdot)$ es la función de activación.

***Definición* 2.11.** *Sean $x^{(k)}$ la muestra k-ésima de entrenamiento de la red MLP y L_n la capa de salida de la red, se define el error cuadrático por*

$$E\left(k\right) = \frac{1}{2}\sum_{j=1}^{n_3}\left(d_j(k) - Y_j^{L_n}(k)\right)^2, \tag{2.98}$$

donde $d_j(k)$ es el valor esperado de la j-ésima neurona de salida para la k-ésima muestra de entrenamiento y $Y_j^{L_n}(k)$ es la salida obtenida en la j-ésima neurona de salida de la red obtenida con la muestra de entrenamiento.

La medición del rendimiento global del algoritmo backpropagation se puede calcular como un promedio del error cuadrático de cada una de las neuronas de salida, técnicamente no es un error medio cuadrático pero se lo notará como tal [34].

***Definición* 2.12.** *El error medio cuadrático para la red MLP con respecto a p muestras de entrenamiento se define como*

$$E_M = \frac{1}{p}\sum_{k=1}^{p} E(k) \tag{2.99}$$

Para desarrollar el algoritmo de propagación hacia atrás se considerarán dos etapas, en la primera se realizará el ajuste de la matriz de pesos $W_{ji}^{(3)}$ de la capa de salida. La segunda parte se refiere a los procedimientos para ajustar las matrices de pesos $W_{ji}^{(2)}$ y $W_{ji}^{(1)}$ asociadas a las capas ocultas.

En la primera etapa el objetivo del entrenamiento es ajustar la matriz de pesos $W_{ji}^{(3)}$ para minimizar el error de las salidas de la red respecto de las salias esperadas. El proceso de ajuste es similar al aplicado en la red ADALINE, por lo tanto, considerando la ecuación (2.98) y empleando la definición de gradiente y la regla de la cadena tenemos

$$\begin{aligned}\nabla E^{(3)} &= \frac{\partial E}{\partial W_{ji}^{(3)}} \\ &= \frac{\partial E}{\partial Y_j^{(3)}} \cdot \frac{\partial Y_j^{(3)}}{\partial I_j^{(3)}} \cdot \frac{\partial I_j^{(3)}}{\partial W_{ji}^{(3)}}\end{aligned} \tag{2.100}$$

A partir de las ecuaciones (2.96), (2.97) y (2.98) se tiene:

$$\frac{\partial I_j^{(3)}}{\partial W_{ji}^{(3)}} = Y_i^{(2)} \tag{2.101}$$

$$\frac{\partial Y_j^{(3)}}{\partial I_j^{(3)}} = g'\left(I_j^{(3)}\right) \tag{2.102}$$

$$\frac{\partial E}{\partial Y_j^{(3)}} = -\left(d_j - Y_j^{(3)}\right), \tag{2.103}$$

donde $g'(\cdot)$ representa la primera derivada de la función de activación.

Reemplazando las ecuaciones anteriores en (2.100) se tiene

$$\frac{\partial E}{\partial W_{ji}^{(3)}} = -\left(d_j - Y_j^{(3)}\right) \cdot g'\left(I_j^{(3)}\right) \cdot Y_i^{(2)}. \tag{2.104}$$

Tomando en cuenta que para minimizar el error debemos tomar la dirección opuesta del gradiente, tenemos

$$\begin{aligned}\Delta W_{ji}^{(3)} &= -\eta \cdot \frac{\partial E}{\partial W_{ji}^{(3)}} \\ &= -\eta \cdot \underbrace{-\left(d_j - Y_j^{(3)}\right) \cdot g'\left(I_j^{(3)}\right)}_{\delta_j^{(3)}} \cdot Y_i^{(2)} \\ &= \eta \cdot \delta_j^{(3)} \cdot Y_i^{(2)}. \end{aligned} \tag{2.105}$$

Expresando (2.105) de forma iterativa

$$W_{ji}^{(3)}(actual) = W_{ji}^{(3)}(previo) + \eta \cdot \delta_j^{(3)} \cdot Y_i^{(2)}. \tag{2.106}$$

Con la aplicación de la expresión (2.106) se ajustan los pesos neuronales de la capa de salida.

El ajuste de los pesos sinápticos de las capas intermedias, dado que no se tiene los valores deseados correspondientes, se los realiza luego de que las neuronas de la capa posterior se hayan ajustado. Tomando nuevamente la red presentada en la Figura (2.11), los pesos de la segunda capa oculta se ajustaran luego de que el ajuste de las neuronas de la capa de salida haya concluido; es en este aspecto ene el que reside la esencia del algoritmo de propagación hacia atrás.

El proceso de entrenamiento para ajustar los pesos de la segunda capa neuronal consiste en ajustar la matriz de pesos $W_{ji}^{(2)}$, tratando de minimizar el error producido entre estas salidas y el error propagado hacia atrás que se origina al ajustar la capa de salida. Considerando la ecuación (2.98) con respecto a la segunda capa y empleando la definición de gradiente:

$$\Delta E^{(2)} = \frac{\partial E}{\partial W_{ji}^{(2)}} = \frac{\partial E}{\partial Y_j^{(2)}} \cdot \frac{\partial Y_j^{(2)}}{\partial I_j^{(2)}} \cdot \frac{\partial I_j^{(2)}}{\partial W_{ji}^{(2)}} \tag{2.107}$$

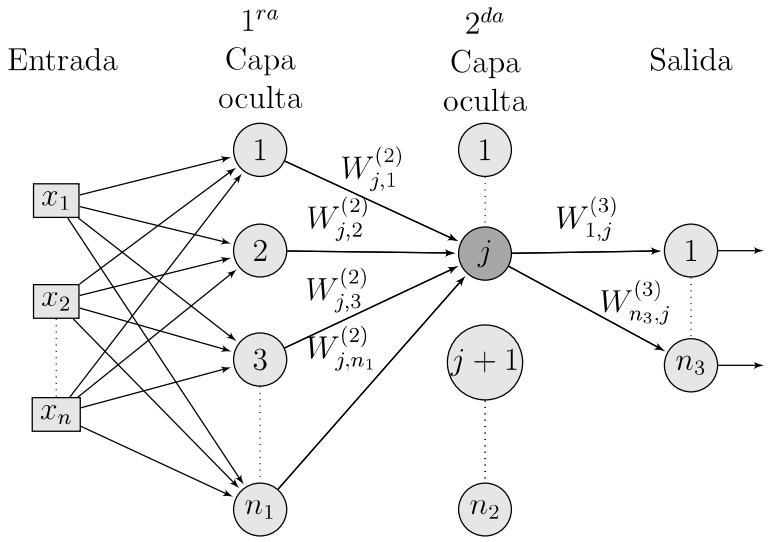
Tomando derivadas parciales se tiene

$$\frac{\partial I_j^{(2)}}{\partial W_{ji}^{(2)}} = Y_i^{(1)} \tag{2.108}$$

$$\frac{\partial Y_j^{(2)}}{\partial I_j^{(2)}} = g'\left(I_j^{(2)}\right) \tag{2.109}$$

Para la derivada de $\frac{\partial E}{\partial Y_j^{(2)}}$ hay que considerar $I_j^{(3)}$ en función de $Y_j^{(2)}$, en la Figura (2.13) se muestra esta representación usando la neurona j de la segunda capa como referencia.

Figura 2.13: Representación de las conexiones neuronales para ajustar los pesos de las neuronas de la segunda capa



Se tiene entonces la derivada:

$$\begin{aligned}
\frac{\partial E}{\partial Y_j^{(2)}} &= \sum_{k=1}^{n_3} \frac{\partial E}{\partial I_k^{(3)}} \cdot \frac{\partial I_k^{(3)}}{\partial Y_j^{(2)}} \\
&= \sum_{k=1}^{n_3} \frac{\partial E}{\partial I_k^{(3)}} \cdot \frac{\partial \left(\sum_{t=1}^{n_3} W_{tj}^{(3)} \cdot Y_j^{(2)} \right)}{\partial Y_j^{(2)}} \\
&= \sum_{k=1}^{n_3} \frac{\partial E}{\partial I_k^{(3)}} \cdot W_{kj}^{(3)} \qquad (2.110) \\
&= -\sum_{k=1}^{n_3} \delta_k^{(3)} \cdot W_{kj}^{(3)} \qquad (2.111)
\end{aligned}$$

Para obtener la ecuación (2.111) se usó la igualdad (2.96) y $\delta_k^{(3)}$ se obtuvo igual que en la ecuación (2.105). Reemplazando (2.108), (2.109) y (2.111) en (2.107) tenemos:

$$\frac{\partial E}{\partial W_{ji}^{(2)}} = -\left(\sum_{k=1}^{n_3} \delta_k^{(3)} \cdot W_{kj}^{(3)} \right) \cdot g'\left(I_j^{(2)} \right) \cdot Y_i^{(1)} \qquad (2.112)$$

Tomando en cuenta que el ajuste de la matriz de pesos $W_{ji}^{(2)}$ se debe realizar en

dirección opuesta al gradiente para minimizar el error, tenemos:

$$\begin{aligned}\Delta W_{ji}^{(2)} &= -\eta \cdot \frac{\partial E}{\partial W_{ji}^{(2)}} \\ &= \eta \left(\sum_{k=1}^{n_3} \delta_k^{(3)} \cdot W_{kj}^{(3)} \right) \cdot g' \left(I_j^{(2)} \right) \cdot Y_i^{(1)} \\ &= \eta \cdot \delta_j^{(2)} \cdot Y_i^{(1)}, \end{aligned} \tag{2.113}$$

convirtiendo la expresión anterior a una forma iterativa:

$$W_{ji}^{(2)}(actual) = W_{ji}^{(2)}(previo) + \eta \cdot \delta_j^{(2)} \cdot Y_i^{(1)} \tag{2.114}$$

Finalmente procedemos a ajustar los pesos de la matriz $W_{ji}^{(1)}$, con lo cual termina el proceso de entrenamiento de la primera capa oculta. Realizando el procedimiento similar que con las anteriores capas, tenemos:

$$\begin{aligned}\Delta E^{(1)} &= \frac{\partial E}{\partial W_{ji}^{(1)}} \\ &= \frac{\partial E}{\partial Y_j^{(1)}} \cdot \frac{\partial Y_j^{(1)}}{\partial I_j^{(1)}} \cdot \frac{\partial I_j^{(1)}}{\partial W_{ji}^{(1)}}. \end{aligned} \tag{2.115}$$

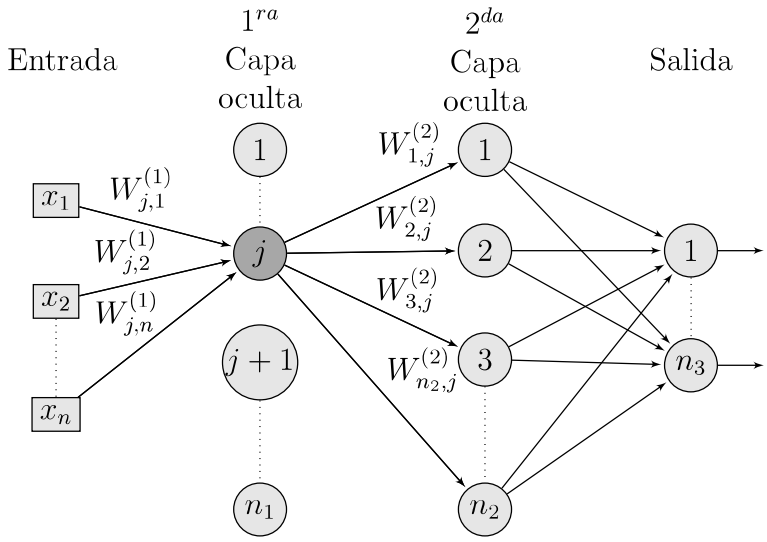
Realizando las derivadas parciales, se obtiene:

$$\frac{\partial I_j^{(1)}}{\partial W_{ji}^{(1)}} = x_i \tag{2.116}$$

$$\frac{\partial Y_j^{(1)}}{\partial I_j^{(1)}} = g' \left(I_j^{(1)} \right) \tag{2.117}$$

En la Figura (2.14) se muestra la interacción de la j-ésima neurona de la primera capa

Figura 2.14: Representación de las conexiones neuronales para ajustar los pesos de las neuronas de la primera capa



oculta, y siguiendo el procedimiento para obtener (2.111), se obtiene:

$$\begin{aligned}\frac{\partial E}{\partial Y_j^{(1)}} &= \sum_{k=1}^{n_2} \frac{\partial E}{\partial I_k^{(2)}} \cdot \frac{\partial I_k^{(2)}}{\partial Y_j^{(1)}} \\ &= \sum_{k=1}^{n_2} \frac{\partial E}{\partial I_k^{(2)}} \cdot \frac{\partial \left(\sum_{t=1}^{n_2} W_{tj}^{(2)} \cdot Y_j^{(1)} \right)}{\partial Y_j^{(1)}} \\ &= \sum_{k=1}^{n_2} \frac{\partial E}{\partial I_k^{(2)}} \cdot W_{kj}^{(2)} \\ &= -\sum_{k=1}^{n_2} \delta_k^{(2)} \cdot W_{kj}^{(2)}. \end{aligned} \tag{2.118}$$

Reemplazando(2.116), (2.117) y (2.118) en (2.115):

$$\frac{\partial E}{\partial W_{ji}^{(1)}} = -\left(\sum_{k=1}^{n_2} \delta_k^{(2)} \cdot W_{kj}^{(2)} \right) \cdot g'\left(I_j^{(1)} \right) \cdot x_i \tag{2.119}$$

Ajustando los pesos de la matriz $W_{ji}^{(1)}$ en dirección contraria al gradiente para mini-

mizar el error, se tiene:

$$\begin{aligned}\Delta W_{ji}^{(1)} &= -\eta \frac{\partial E}{\partial W_{ji}^{(1)}} \\ &= \eta \underbrace{\left(-\sum_{k=1}^{n_2} \delta_k^{(2)} \cdot W_{kj}^{(2)} \right) \cdot g'\left(I_j^{(1)} \right)}_{\delta_j^{(1)}} \cdot x_i \\ &= \eta \cdot \delta_j^{(1)} \cdot x_i \end{aligned} \tag{2.120}$$

Tenemos el proceso iterativo

$$W_{ji}^{(1)}(actual) = W_{ji}^{(1)}(previo) + \eta \cdot \delta_j^{(1)} \cdot x_i \tag{2.121}$$

Se dirá que el algoritmo convergen cuando la diferencia entre el error medio cuadrático de dos iteraciones sucesivas es suficientemente pequeño, es decir

$$\left| E_M^{t+1} - E_M^t \right| \leq \varepsilon, \tag{2.122}$$

donde t es el contador de iteraciones realizadas y ε es la precisión de convergencia definida.

El algoritmo backprop de la fase de entrenamiento de la red MLP se presenta en el Algoritmo 6.

2.2.6. Versiones optimizadas del algoritmo backprop

Algoritmo Resilient-Propagation (Rprop)

Este algoritmo considera la variación del signo del gradiente de la función de error, de la ecuación 2.98, en lugar de considerar las variaciones de su magnitud. En este caso la tasa de aprendizaje se hace dinámica, dado que si en dos pasos consecutivos el signo del gradiente es el mismo entonces la tasa de aprendizaje se incrementa. En otro caso, cuando los signos son distintos, se tiene una reducción de la tasa de aprendizaje.

Algoritmo 6: Algoritmo MLP: Entrenamiento

Obtener el conjunto de entrenamiento $\left\{\mathbf{x}^{(k)}\right\}$
Asociar el vector de salidas $\left\{\mathbf{d}^{(k)}\right\}$ con cada muestra de entrenamiento
Inicializar $W_{ji}^{(1)}$, $W_{ji}^{(2)}$ y $W_{ji}^{(3)}$ con valores aleatorios pequeños
Especificar la tasa de aprendizaje η y la precisión ε
Inicializar un contador $c \leftarrow 0$
repetir
    $E_M^c \leftarrow E_M$
    para $\left\{\mathbf{x}^{(k)}, \mathbf{d}^{(k)}\right\}$ **hacer**
        Pasos hacia adelante:
        Obtener $l_j^{(1)}$ y $Y_j^{(1)}$
        Obtener $l_j^{(2)}$ y $Y_j^{(2)}$
        Obtener $l_j^{(3)}$ y $Y_j^{(3)}$
        Pasos hacia atrás:
        Determinar $\delta_j^{(3)}$
        Ajustar $W_{ji}^{(3)}$
        Determinar $\delta_j^{(2)}$
        Ajustar $W_{ji}^{(2)}$
        Determinar $\delta_j^{(1)}$
        Ajustar $W_{ji}^{(1)}$
    Obtener $Y_j^{(3)}$ ajustada
    $E_M^{c+1} \leftarrow E_M^c$
    $c \leftarrow c + 1$
hasta que $\left|E_M^{c+1} - E_M^c\right| \leq \varepsilon$

Matemáticamente el cambio de signo del gradiente se expresa de la siguiente manera:

$$\Lambda_{ji}^{(L)}(t) = \begin{cases} \eta^+ \cdot \Lambda_{ji}^{(L)}(t-1) & \text{, si } \dfrac{\partial E(t-1)}{\partial W_{ji}} \cdot \dfrac{\partial E(t)}{\partial W_{ji}} > 0 \\ \eta^- \cdot \Lambda_{ji}^{(L)}(t-1) & \text{, si } \dfrac{\partial E(t-1)}{\partial W_{ji}} \cdot \dfrac{\partial E(t)}{\partial W_{ji}} < 0 \\ \Lambda_{ji}^{(L)}(t-1) & \text{, caso contrario,} \end{cases} \tag{2.123}$$

donde $\Lambda_{ji}^{(L)}$ es la tasa de aprendizaje individual asociada a cada $W_{ji}^{(L)}$, donde $0 < \eta^- < 1$ y $1 < \eta^+$ [34].

Algoritmo Rprop con retroceso de peso (Rprop+)

Este algoritmo[2] implementa una modificación en las actualizaciones de los pesos ΔW_{ij} considerando dos casos: si el signo de la derivada parcial no cambia se ejecuta la actualización de peso regular:

$$\text{si } \frac{\partial E(t-1)}{\partial W_{ji}} \cdot \frac{\partial E(t)}{\partial W_{ji}} \geq 0 \text{ entonces } \Delta W_{ij}(t) = -sign\left(\frac{\partial E(t)}{\partial W_{ji}}\right) \cdot \Lambda_{ji}(t). \tag{2.124}$$

Pero en caso de cambio de signo de la derivada parcial se revierte la actualización del peso anterior:

$$\text{si } \frac{\partial E(t-1)}{\partial W_{ji}} \cdot \frac{\partial E(t)}{\partial W_{ji}} < 0 \text{ entonces } \Delta W_{ij}(t) = -\Delta W_{ij}(t-1) \text{ y } \frac{\partial E(t)}{\partial W_{ji}} = 0. \tag{2.125}$$

Algoritmo Rprop sin retroceso de peso (Rprop-)

En esta versión[3] se omite el retroceso de peso (*Weight-Backtracking*) y se utiliza solo el lado derecho de la ecuación 2.124 en cualquier caso, por lo tanto, no es necesario almacenar las actualizaciones de los pesos anteriores.

[2]El término *Rprop con retroceso de peso* es una traducción por parte del autor, el término original es *Rprop with Weight-Backtracking*. Por lo que se puede encontrar al algoritmo con otros nombres en la literatura.

[3]El término *Rprop sin retroceso de peso* es una traducción por parte del autor, el término original es *Rprop without Weight-Backtracking*. Por lo que se puede encontrar al algoritmo con otros nombres en la literatura.

Capítulo 3

Análisis de los datos y clusterización

3.1. Descripción de la base de datos

La información se obtuvo del Operativo de Rendimientos Objetivos de Maíz Duro Seco que tiene en marcha desde el MAGAP; los datos corresponden desde el ciclo de verano de 2014 al ciclo invierno de 2017, los mismos que cumplen los criterios de validación determinados en el Manual del Investigador [6]. Como se describió en la sección 1.3, los datos son levantados de las seis provincias productoras de maíz más representativas: Los Ríos, Guayas, Loja, Manabí El Oro y Santa Elena.

Se dispone de 1.220 encuestas, de los cuales 578 corresponden al ciclo de invierno y 642 al ciclo de verano, se considera esta división en el tratamiento de los datos puesto que se desarrollarán dos modelos; en el siguiente capítulo se detallará cada modelo.

3.1.1. Selección de las variables

La elección de las variables independientes se realizó en base al criterio de los expertos de la CGSIN que determinaron las siguientes: la temperatura promedio del ciclo (°C), la precipitación[1] acumulada del ciclo (mm), la fertilización aplicada (qq/ha), variedad/híbrido, paquete tecnológico, superficie sembrada del cultivo [30]. Adicional a estas variables se consideran otras que podrían influir en el rendimiento: capacitación, miembro de una asociación, cultivo solo o asociado, cantidad de material vegetativo, preparación del suelo,

[1]La precipitación se mide como la altura de agua caída sobre una superficie plana, tal altura se expresa en milímetros mm. Para estudios de irrigación se tiene la equivalencia: $1mm = \frac{10m^3}{ha}$[42].

riego, gasto en producción, mazorcas en diez metros lineales, cosecha, provincia, cantón.

Los campos de la temperatura promedio y precipitación acumulada se obtuvieron de las estaciones del INAMHI y se asignaron de acuerdo a la ubicación geográfica y fecha de siembra de cada cultivo, considerando que el ciclo de cultivo en Los Ríos, Manabí, Guayas, Santa Elena y El Oro es de cuatro meses aproximadamente y en zonas altas como Loja el ciclo es de cinco meses aproximadamente [7], [30].

A continuación se describe cada una de las variables

- **Temperatura promedio del ciclo:** es una variable cuantitativa que representa la temperatura promedio desde la fecha de siembra hasta la fecha de cosecha. Temperaturas cálidas durante la siembra y últimas semanas del ciclo ayudan a la germinazión de la semilla y maduración del grano. "Sin embrago, temperaturas superiores a los 35°C tiene efectos negativos sobre la fotosíntesis, la traslocación, la fertilidad de las florecillas, el éxito de la polinización y otros aspectos del metabolismo que reducen la productividad de la planta"[30]. Bajo esta consideración, se espera que la relación entre el rendimiento y la temperatura siga una curva en U invertida. En el cuadro 3.1 se presentan los valores mínimo y máximo de la temperatura promedio de cada ciclo de cultivo.

Cuadro 3.1: Temperatura promedio ciclo

		Temperatura	
		Mínima	Máxima
Ciclo	Invierno	26,163°C	27,472°C
	Verano	24,627°C	26,446°C

- **Precipitación acumulada del ciclo:** es una variable cuantitativa que representa la precipitación acumulada del ciclo. La falta de agua reduce la formación de granos y de mazorcas, así como un llenado parcial y lento de las mazorcas. Por otro lado, cuando se producen inundaciones las plantas disminuyen el crecimiento de las raíces, la absorción de nutrientes y la capacidad reproductiva de la planta. Por tal motivo se espera una relación en U invertida entre el rendimiento y la precipitación [30].

  En el cuadro 3.2 se presenta las precipitaciones mínima y máxima del ciclo de invierno y de verano; como se puede observar, el ciclo de invierno tienen una precipitación máxima que supera lo recomendado por la CGSIN, que es del 2 000 mm [30].

Cuadro 3.2: Precipitación acumulada ciclo

		Precipitación	
		Mínima	Máxima
Ciclo	Invierno	58,60mm	2380,10mm
	Verano	0,90mm	565,40mm

- **Fertilización aplicada:** es una variable cuantitativa que representa la cantidad en quintales por hectárea (qq/ha) de fertilizante usado. Los fertilizantes más comunes son UREA, MOP, DAP, y Sulpomag; cada uno de estos están compuestos de Nitrógeno, Fósforo y Potasio. Por lo expuesto, se registra la fertilización como tres variables separadas, las cuales representan la cantidad en quintales por hectárea usados de cada uno de los componentes [6].

Cuadro 3.3: Fertilización aplicada

		Nitrógeno		Fósforo		Potasio	
		Mínima	Máxima	Mínima	Máxima	Mínima	Máxima
Ciclo	Invierno	0,46 qq/ha	4,14 qq/ha	0 qq/ha	1,5 qq/ha	0 qq/ha	1,48 qq/ha
	Verano	0 qq/ha	5,28 qq/ha	0 qq/ha	1,84 qq/ha	0 qq/ha	1,6 qq/ha

Las fertilización aporta los principales macronutrientes que no están presentes en los suelos tropicales y que a falta de ellos se reducen el crecimiento de las hojas, las mazorcas y los granos. No obstante, su uso en cantidades indebidas producen toxicidad en el suelo lo cual limita el crecimiento de esta gramínea; por lo tanto se espera una relación de incrementos decrecientes entre el rendimiento y la cantidad de fertilizante usado. Se recomienda de 5 qq/ha a 7 qq/ha de nitrógeno, de 2 qq/ha a 3 qq/ha de potasio y de 2 qq/ha a 3 qq/ha de fósforo. En el cuadro 3.3 se presentan las cantidades mínimas y máximas de cada compuesto; como se observa, ninguno supera los valores límite recomendados [30].

- **Variedad/híbrido:** es una variable nominal que representa a la variedad de maíz que usó el agricultor en su campo. El híbrido más utilizado es el Trueno NB 7443, que posee las siguientes características: alta resistencia a las plagas, alta tolerancia al volcamiento[2] y uniformidad en la altura de la mazorca [3]. Otras de las variedades más usadas son: Dekalab 7088, Insignia 105, Somma y Triunfo.

[2]El *Volcamiento del maíz* es una enfermedad que según el INIAP presenta la sintomatología ": flacidez del cogollo con marchitamiento completo de los tejidos foliares; secamiento y muerte de la planta."[21]

- **Paquete tecnológico:** es una variable dummy que toma el valor de uno si el agricultor recibió el Kit del programa Plan Semilla de Alto Rendimiento que es subsidiado por el gobierno; el programa permite el acceso a insumos como fertilizantes, semillas y agroquímicos de mejor calidad [3].

Cuadro 3.4: Agricultores que recibieron el Kit de beneficio

		Sí recibió	No recibió
Ciclo	Invierno	193 (33 %)	385 (67 %)
	Verano	232 (36 %)	410 (64 %)

Más del 60 % de los agricultores consultados afirman que no recibieron el Kit, esto se puede observas en los resultados presentados en el cuadro 3.4.

- **Superficie sembrada del cultivo:** es una variable cuantitativa que representa a la superficie en hectáreas que fueron destinados para el cultivo de maíz. Se suele usar esta variable como *proxy*[3] del nivel de tecnificación, dado que existe una fuerte correlación entre el tamaño del agricultor y su capacidad de inversión en infraestructura y mecanización, así como el acceso a insumos de mejor calidad y mayor conocimiento técnico del manejo de los cultivos [30].

- **Capacitación:** es una variable dummy que toma el valor de uno si el agricultor recibió capacitación en el último año.

Cuadro 3.5: Capacitación

		Sí recibió	Rendimiento promedio	No recibió	Rendimiento promedio
Ciclo	Invierno	279 (48 %)	5,41 Tm/ha	299 (52 %)	5,08 Tm/ha
	Verano	384 (60 %)	6,04 Tm/ha	248 (40 %)	5,98 Tm/ha

Analizando el cuadro 3.5 se puede constatar que en el ciclo de invierno existe una diferencia de dos puntos porcentuales entre los que recibieron capacitación (48 %) y los que no (52 %). A diferencia, en el ciclo de verano es superior la cantidad de agricultores que recibieron capacitación pues representan el 60 % del total. Como es de esperar, el rendimiento de los que sí recibieron capacitación es mejor en ambos ciclos.

- **Miembro de una asociación:** es una variable dummy que toma el valor de uno si el agricultor es miembro de una asociación. En el cuadro 3.6 se presenta la cantidad

[3]Una variable *proxy* es aquella que se usa como aproximación de otra variable de estudio de la cual no se tiene acceso.

Cuadro 3.6: Agricultores miembros de asociaciones

		Asociado	Rendimiento promedio	No asociado	Rendimiento promedio
Ciclo	Invierno	164 (28 %)	5,46 Tm/ha	414 (72 %)	5,15 Tm/ha
	Verano	323 (50,3 %)	5,84 Tm/ha	319 (49,7 %)	6,22 Tm/ha

de agricultores asociados y no asociados para los dos ciclos de producción. En el ciclo de invierno, como es de esperar, es mayor el número de agricultores no asociados, esto debido a que en el invierno siembran todos aquellos agricultores que poseen un terreno mas o menos apto para el cultivo. Algo muy distinto sucede en el ciclo de verano, donde el porcentaje de participación de los asociados y no asociados es prácticamente del 50 %. Cabe destacar que los que siembran en verano son aquellos agricultores que cuentan con una infraestructura de riego; y, dado que esto conlleva una mayor inversión, se relaciona también un manejo más tecnificado y el uso de insumos de alto rendimiento; esto explicaría el hecho de que los agricultores no asociados en el verano presenten un mayor rendimiento por hectárea [30].

- **Cultivo solo o asociado:** es una variable dummy que toma el valor de uno cuando el agricultor informa que el cultivo de maíz es sembrado junto con otro producto.

- **Cantidad de material vegetativo:** es una variable cuantitativa que representa la cantidad de semilla utilizada por hectárea.

- **Preparación del suelo:** es una variable dummy que toma el valor de uno cuando el agricultor informa que realizó una labor de preparación del suelo antes de la siembra: limpieza del suelo, abonado, entre otros.

- **Riego:** es una variable dummy que toma el valor de uno cuando el agricultor informa que posee un sistema de riego artificial, es decir, debe tener cierta estructura que haga posible el riego [6].

- **Gasto en producción:** es una variable cuantitativa que representa el costo total por hectárea del cultivo que el agricultor gastó. Se considera al menos la preparación del suelo, semilla, fertilizantes, mano de obra y maquinaria [6].

- **Mazorcas en diez metros lineales:** es una variable cuantitativa que representa la cantidad de mazorcas obtenidas en diez metros lineales de una hilera del cultivo. El proceso para la obtención de este dato se detalla en al Manual del Investigador [6].

- **Cosecha:** es una variable dummy que toma el valor de uno cuando el agricultor informa que para la cosecha del maíz se utilizó maquinaria

- **Provincia:** es una variable nominal que representa la provincia donde se encuentra el cultivo. En el Operativo de Rendimientos Objetivos se tomó información solo de las principales provincias productoras: Los Ríos, Guayas, Manabí, Loja, El Oro y Santa Elena [3].

- **Cantón:** es una variable nominal que representa el cantón donde se encuentra el cultivo.

La notación usada para los modelos a desarrollar seá la presentada en en cuadro 3.7.

Cuadro 3.7: Notación de las variables obtenidas de la base de datos

Notación	Variables	Notación	Variable
T_P	Temperatura promedio del ciclo	P_A	Precipitación acumulada del ciclo
QQ_N	Quintales de nitrógeno usados por hectáreas	QQ_F	Quintales de fósforo usados por hectárea
QQ_P	Quintales de potasio usados por hectáreas	K_B	Paquete tecnológico
S_C	Superficie sembrada de cultivo	R_C	Capacitación
M_A	Miembro de una asociación	C_S	Cultivo solo o asociado
C_M	Cantidad de material vegetativo	P_S	Preparación del suelo
T_R	Tiene riego	G_P	Gasto en producción
N_M	Mazorcas en diez metros lineales	C1	Cosecha mecanizada
C0	Cosecha no mecanizada		

3.1.2. Clusterización por provincias, cantones y variedad

En el estudio del rendimiento del maíz duro seco se esperaría que la ubicación geográfica sea un factor influyente. Una forma de analizar si la ubicación geográfica, ya sea provincial o cantonal, es un atributo que determina un aumento o disminución del rendimiento del cultivo es mediante es mediante el uso de variables artificiales de escala nominal, que toman el valor de cero o uno en ausencia o presencia de dicho atributo respectivamente; para esto, cada provincia o cantón estará representado por el promedio del rendimiento de los agricultores de esa zona, y estos serán agrupados mediante el algoritmo *k-means* [4] [12].

[4]El agrupamiento de datos (o *clustering*) consiste en crear grupos de objetos de tal manera que los elementos dentro de cada uno de los grupos sean similares respecto de cierta característica, y entre elementos de diferentes grupos sean distintos. Uno de los métodos más simples y clásicos es el algoritmo *k-means*.

Un factor que es considerado en el rendimiento del cultivo de maíz es la variedad sembrada.

Clustericazión por provincias

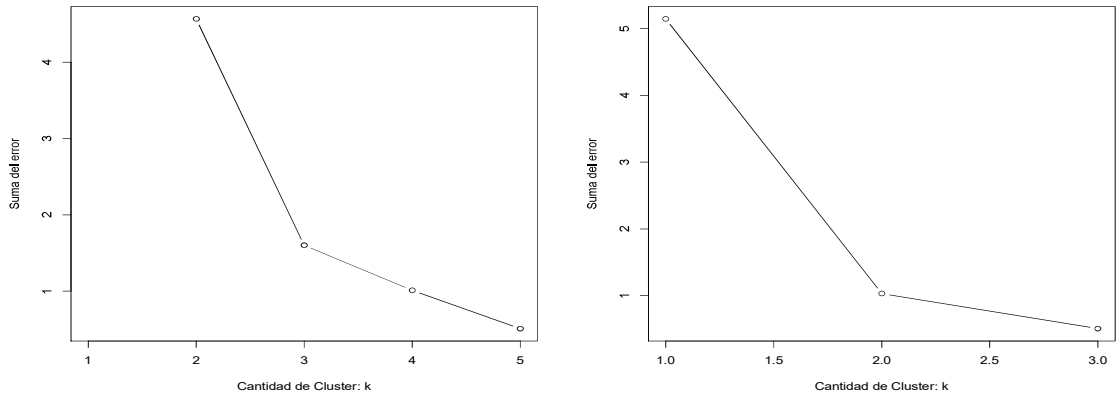
Se calculó para cada provincia el rendimiento promedio, esto tanto para las provinicas que presentaron producción en invierno y en verano. En la figura 3.1 se presenta el rendimiento promedio por provincia y por ciclo de producción; podemos notar que en el ciclo de verano no existen registros de dos provincias a diferencia del ciclo invierno: Loja y El Oro.

Figura 3.1: Rendimiento promedio por provincia y ciclo

(a) Rendimiento promedio para el ciclo invierno (b) Rendimiento promedio para el ciclo verano

Se aplicó el algoritmo *k-means* a los grupos de provincias de invierno y verano, y se graficó los suma de errores por cada clúster respecto del número k de clústeres elegidos; los resultados se presentan en la figura 3.2. Para determinar el número k óptimo se considera el clúster en el cual se forma una *rodilla* en el gráfico de error contra los valores de k [13]. Con estas consideraciones, se tiene que el número óptimo de clústeres para el ciclo de invierno es $k = 3$ y para el ciclo de verano es $k = 2$.

Figura 3.2: Suma del error interclúster provincial para los distintos valores de k elegidos



(a) Clusterización del ciclo invierno

(b) Clusterización del ciclo verano

Aplicando el algoritmo para cada uno de los k óptimos obtenemos la clusterización por provincias que se presenta en el cuadro 3.8.

Cuadro 3.8: Clusterización por provincias

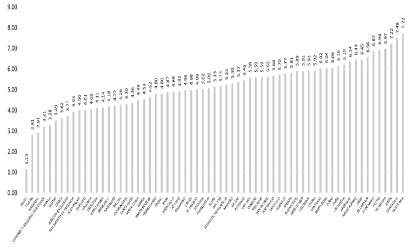
Invierno		Verano	
Notación del Clúster	Provincia	Notación del Clúster	Provincia
PI1	El Oro, Loja	PV1	Manabí, Santa Elena
PI2	Guayas, Santa Elena	PV2	Los Ríos, Guayas
PI3	Los Ríos, Manabí		

Clustericazión por cantón

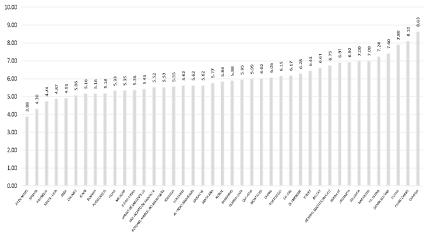
Por cantón y ciclo se calculó el rendimiento promedio de maíz duro seco. Los resultados obtenidos se presentan en la figura 3.3. Se puede observar que los rendimientos en el ciclo de verano son más homogéneos por lo que se eperaría que el número de clústeres en verano sea menor que en invierno.

Se ejecutó el algoritmo *k-means* a los grupos de cantones de invierno y verano, y se graficó los suma de errores por cada clúster respecto del número k de clústeres elegidos y los resultados se presentan en la figura 3.4. Se determinó el número k óptimo de clústeres: $k = 4$ para invierno y $k = 3$ para verano. Los resultadps de la aplicación del algoritmo *k-means* con los k óptimos se presentan en el cuadro 3.9.

Figura 3.3: Rendimiento promedio por cantón y por ciclo

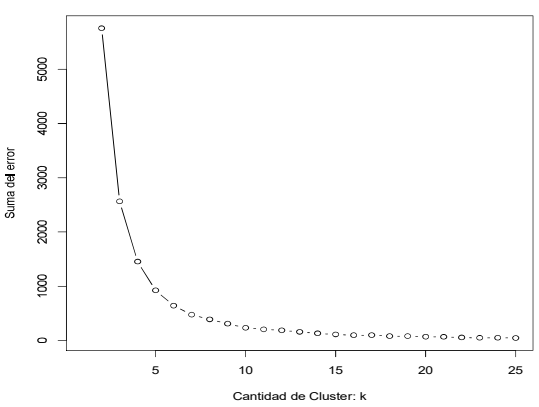


(a) Clusterización del ciclo invierno

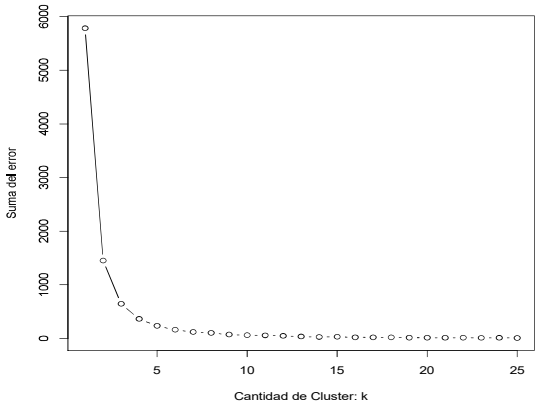


(b) Clusterización del ciclo invierno

Figura 3.4: Suma del error interclúster cantonal para los distintos valores de k elegidos



(a) Clusterización del ciclo invierno



(b) Clusterización del ciclo verano

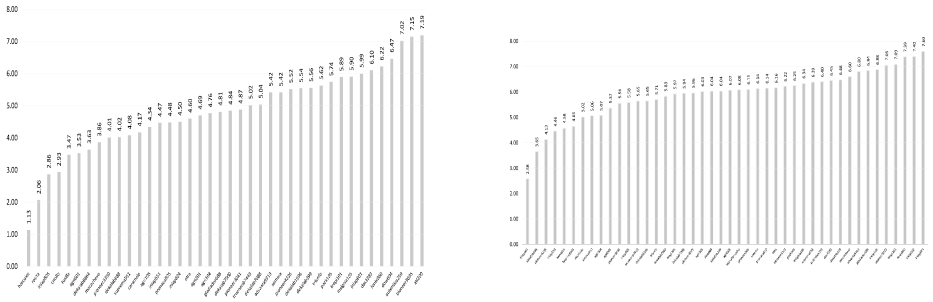
Cuadro 3.9: Clúster por Cantones

Invierno		Verano	
Clúster	Cantón	Clúster	Cantón
CI1	Ventanas, Santa rosa, Piñas, Paján, Urdaneta, Arenillas, Flavio alfaro, Jama, El carmen, 24 de mayo, Paltas, Pichincha, Junín, Guayaquil, Santa ana	CV1	San Jacinto de Yaguachi, Coronel Marcelino Maridueña, Tosagua, Ventanas, Alfredo baquerizo, Mocache, Santa Ana, Nobol, Babahoyo, Quinsaloma, Quevedo, Montalvo, Chone, Portoviejo
CI2	Daule, Durán, Naranjal, Coronel marcelino maridueña, Manta, Salitre, Celica, Alfredo baquerizo, San Jacinto de yaguachi, El empalme, Quevedo, Valencia, Palestina, Isidro ayora, Marcabelí, Naranjito, Balzar	CV2	Salitre, El Empalme, Vinces, Balzar, General Elizalde (Bucay), Buena Fe, Urdaneta, Valencia, Naranjito, Palestina, Simón Bolívar, Durán, Pedro Carbo, Olmedo
CI3	Lomas de sargentillo, Milagro, Balsas, Puyango, Colimes, Olmedo, Mocache, San vicente, Portoviejo, Zapotillo, Buena Fe, Jipijapa, Rocafuerte, Montecristi, Palenque, Chone	CV3	24 de Mayo, Jipijapa, Palenque, Santa Lucía, Baba, Colimes, Junín, Bolívar, Puebloviejo, Paján, Milagro, Isidro Ayora, Lomas de Sargentillo
CI4	Quinsaloma, Santa lucía, Santa elena, Nobol, Simón Bolívar, Pedro carbo, Pindal, Baba, Montalvo, Las lajas, Babahoyo, Vinces, El triunfo, Tosagua, Puebloviejo, Sucre, Bolivar		

Clusterización por variedad

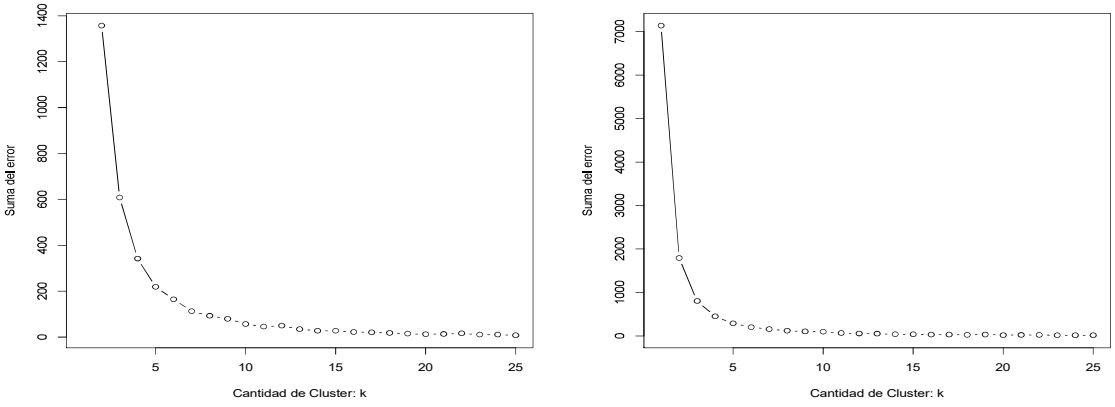
Considerando las variedades de maíz sembradas, se calculó el rendimiento promedio tanto para el ciclo de invierno como para el de verano, en la figura 3.5 se presentan los resultados. Realizando el procedimiento de clusterización aplicado a provincias y cantones obtenemos que el número óptimo de clústeres por ciclo son: $k = 4$ para invierno y $k = 3$ para verano, por observación de la figura 3.6.

Figura 3.5: Rendimiento promedio por variedad y por ciclo



(a) Clusterización del ciclo invierno (b) Clusterización del ciclo invierno

Figura 3.6: Suma del error interclúster de variedad para los distintos valores de k elegidos



(a) Clusterización del ciclo invierno (b) Clusterización del ciclo verano

La clusterizaión por variedad obtenida aplicando del algoritmo *k-means* con los k óptimos se presentan en el cuadro 3.10.

Cuadro 3.10: Clúster por variedad

Invierno		Verano	
Clúster	Variedad	Clúster	Variedad
VI1	Hercules, Rosita, Iniap505, Criollo, Tusilla, Agri601, Dekalab8864, Mocachevo, Pioneer3350, Dekalab688	VV1	Iniap601, Dekalab688, Pioneer4226, Iniap551, Hercules, Brasilia8501, Reciclada, Senaca810, Agri104, Agri201, Pioneer3035, Iniap505, Truenonb7443, Dekalab1596, Triunfo
VI2	Suprema551, Caramelo, Agri105, Iniap551, Pronaca505, Iniap824, Otra, Agri201, Agri104, Gladiador688, Dekalab7500	VV2	Insignia105, Suprema551, Autentica259, Das3383, Dekalab399, Mocachevo, Advanta9313, Gladiador688, Iniap105, Pioneer3073, Tropi101, Fenix860, Iniap602, Iniap824
VI3	Tropi101, Insignia105, Iniap601, Das3383, Fenix860, Dow604, Autentica259, Pioneer3035, Atl200	VV3	Dekalab7500, Tropi105, Dekalab7088, Pioneer3075, Agri105, Dow604, Dekalab299, Agri601, Manabimocho, Pronaca505, Somma, Pronaca810, Otra, Pioneer3041, Pack105
VI4	Pioneer3041, Truenonb7443, Dekalab7088, Advanta9313, Somma, Pioneer4226, Dekalab1596, Dekalab399, Triunfo, Pack105		

Capítulo 4

Desarrollo de un Modelo de Regresión Lineal Múltiple y un Modelo de Redes Neuronales Artificiales

En este capítulo se presentan los modelos obtenidos tanto de regresión lineal múltiple como de redes neuronales artificiales para el ciclo de invierno y el de verano, ambos modelos de realizando usando el software estadístico R. Se determinan las variables que influyen de manera significativa al rendimiento de maíz por hectárea y se validan los modelos de regresión.

Se calcula las cotas inferior y superior del número de neuronas en la capa oculta de las redes neuronales, y se desarrolla cada una de estas arquitecturas con los diferentes algoritmos y se selecciona la arquitectura que mejores estadísticos de bondad de ajuste genere. Finalmente, se comparan los resultados obtenidos en ambos modelos.

4.1. Desarrollo de un modelo de regresión lineal múltiple

En la estimación del rendimiento de maíz mediante regresión lineal es común usar como variables regresoras las condiciones climáticas y cantidad de fertilizantes utilizados en el ciclo de cultivo. Matsumura K., *et al*[24] realiza una estimación del rendimiento anual

de maíz en Jilin (China) usando como variables predictoras la temperatura promedio del ciclo, la precipitación acumulada, la cantidad de potasio, nitrógeno y fósforo anual usado; los métodos que usó fueron regresión lineal múltiple y redes neuronales artificiales, a diferencia de lo que se plantea en el presente trabajo, Matsumura utiliza datos anuales desde 1961 a 2004 sin la distinción de ciclos de siembra y cosecha.

Predicciones usando regresión lineal múltiple (RLM) y redes neuronales artificiales (ANN) con datos obtenidos de intervalos de tiempo más cortos y considerando la distinción entre ciclos de invierno y verano se los puede ver en los trabajos de Bilgili [2] y Niazian *et al* [32], a pesar de no estimar el rendimiento de maíz (sino temperatura del suelo y rendimiento de la semilla de Ajowan, respectivamente) dan una guía de tratamiento y comparación de estos dos métodos.

En el Ecuador, el MAGAP a través de la CGSIN determinó un modelo de regresión lineal múltiple para estimar la producción anual de maíz usando como variables predictoras el precio del quintal de maíz al productor, la importación de fertilizantes, la importación de semillas y una variable *dummy* para eventos atípicos como sequías; consideraron datos anuales entre los años 2000 a 2013. Lo que es importante notar de este modelo y que será asumido en el presente proyecto, es que no se considera la constante en el modelo de RLM, bajo el supuesto de que si no existieran las variables independientes seleccionadas, no existiría producción de maíz, lo que implica que el valor mínimo de producción es cero [45].

Con datos del Operativo de Rendimientos Objetivos de los ciclos de invierno y verano del 2014 la CGSIN determinó que existe una diferencia estructural (mediante la prueba de Chow) en el comportamiento del rendimiento en cada uno de los dos ciclos. Dado este resultado, plantean dos regresiones, una para cada ciclo, con el fin de obtener un mejor ajuste del que se lograría si se planteara un único modelo para los dos ciclos en conjunto. Es de esperar que exista esta diferencia estructural, debido a que las condiciones climatológicas son distintas en cada una de las estaciones y afectan directamente a cualquier cultivo agrícola [30].

Para una mejor comparación de los modelos de ANN y RLM se normaliza los datos de entrada, es decir, los datos correspondientes de las variables independientes fueron normalizadas usando la ecuación (2.84) para el intervalo $[0, 1]$[32].

4.1.1. Modelo de RLM para el ciclo invierno

Para ciclo de invierno se dispone de 578 observaciones, de los cuales se toman 421 registros que corresponden a los dos primeros ciclos de invierno (representan el 73 % del total de datos de invierno) para el proceso de entrenamiento y que serán usados para determinar el modelo de RLM para el ciclo de invierno y para el proceso de entrenamiento de las redes neuronales. En 27 % restante serán usados para el test de los modelos.

Estimación de los parámetros del modelo

La variable precipitación acumulada ciclo excede el valor recomendado, como se puede observar en el cuadro 3.2. En concreto son 45 casos en los que la precipitación acumulada es superior al límite establecido por la CGSIN que es de 2,000mm. En primera instancia se podría considerar incluir un polinomio cuadrático en el modelo para evidenciar si existe una relación decreciente entre el rendimiento y la precipitación, no obstante, si la variable P_A no es significativa al 95 % de confianza no es necesario incluir su polinomio [49]. La temperatura en este ciclo no supera lo recomendado como se observa en el cuadro 3.2, por lo tanto no es necesario incluir un polinomio de esta variable.

La fertilización en quintales de fósforo, nitrógeno y potasio por hectárea tampoco supera la cantidad recomendada por la CGSIN, tal como se puede observar en el cuadro 3.3, por lo tanto, no es necesario incluir polinomios de estas variables en el modelo.

Adicional a las variables obtenidas de la base de datos y las generadas por clusterización se consideró las variables generadas por la interacción entre los clústeres de variedad con las variables de fertilización QQ_N, QQ_F y QQ_P; y, con la variable de número de mazorcas N_M. Considerando estas variables se realizó regresión por pasos[1] para determinar las mejores variables predictivas, seleccionando las variables con un nivel de significancia del 5 %. En el cuadro 4.1 se presenta los resultados del modelo obtenido, donde se aprecia que la variable de interacción, notaad vr_i1_qqn, entre el clúster CI1 con la variable de fertilización QQ_N es significativa.

[1]La regresión por pasos es un procedimiento para seleccionar de entre un número grande de posibles variables predictoras las que sean estadísticamente significativas para el modelo. Se parte de un modelo con solo la constante y se van añadiendo las variables que sean significativas a partir del estadístico t; en cada paso también se van comparando la contribución significativa de cada variable y se elimina la de menor contribución de acuerdo a un nivel especificado del valor absoluto del estadístico t [4], [10].

Cuadro 4.1: Modelo de RLM de invierno

Coeficientes	Estimación	Error Std.	valor t	Pr(> |t|)
K_B	0,05112	0,01379	3,706	0,0002 ***
N_M	0,49046	0,03492	14,046	<2e-16 ***
QQ_N	0,08610	0,03188	2,701	0,007199 **
CI1	0,05963	0,01700	3,508	0,000502 ***
CI2	-0,07924	0,01757	-4,510	8,47e-06 ***
VI1	-0,08555	0,03381	-2,530	0,011766 *
vr_i1_qqn	0,30471	0,04596	6,629	1,05e-10 ***
Observaciones	421			
R^2	0,9441			
R^2 Adjustado	0,9432			
Error Std. Residual	0,128 (df = 414)			
Nota:	*p<0,05; **p<0,01, ***p<0,001			

Validación del modelo

- **Prueba RESET:** Para asegurarnos que el modelo no omitió algunas no linealidades importantes, como es el caso de la temperatura, precipitación y fertilización, se realizó la prueba de error de especificación $RESET$[2]. Los resultados de las pruebas para el modelo de invierno se presenta en el cuadro 4.2, al tener los p-valores superiores a 0,05 no se rechaza el hecho de que la forma funcional del modelo es correcta.

Cuadro 4.2: Prueba RESET del modelo de RLM de invierno

Potencia	P-valor
$\hat{y}^2$	0,3921
$\hat{y}^3$	0,2151
$\hat{y}^2, \hat{y}^3$	0,1204

- **Multicolinealidad:** Podemos ver en la Figura 4.1 que no existe pareja de varia-

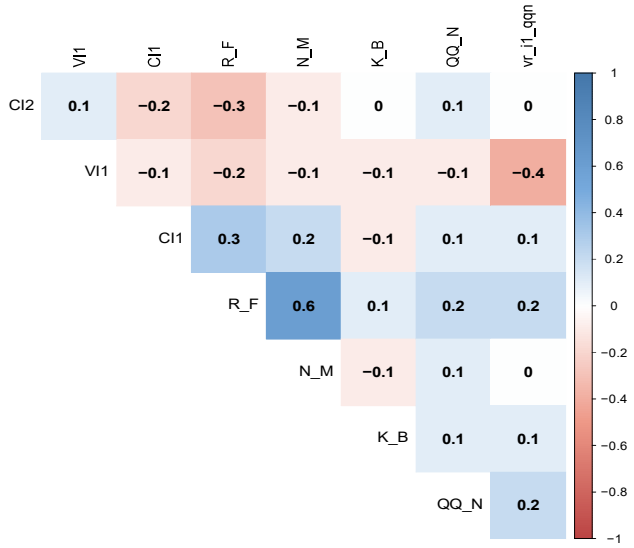
[2]La prueba RESET detecta los tipos generales de especificación incorrecta de la forma funcional; el procedimiento consiste en agregar a la ecuación 2.1 polinomios en los valores ajustados por mínimos cuadrados ordinarios, donde los términos cuadrados y cúbicos suelen ser los más comunes en la mayoría de los casos:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \delta_1 \hat{y}^2 + \delta_2 \hat{y}^3 + w. \tag{4.1}$$

Bajo la hipótesis nula que 2.1 esté correctamente especificada, de manera que RESET es el estadístico F para probar $H_0 : \delta_1 = 0, \delta_2 = 0$ en el modelo 4.1; donde F es aproximadamente $F_{2,n-k-3}$ en muestras grandes.

bles que estén altamente correlacionadas, dado que no existen valores superiores o cercanos a 0.9. El factor de inflación de la varianza no se aplica dado que el modelo no considera la constante.

Figura 4.1: Matriz de correlación del modelo de invierno

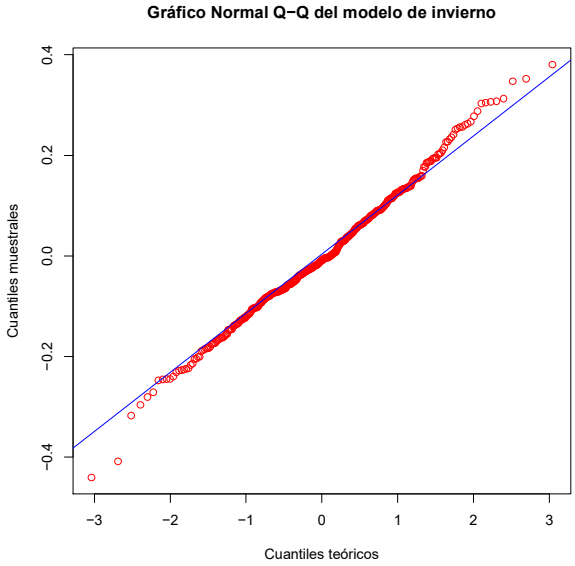


- **Normalidad:** Realizamos primero un análisis gráfico mediante el gráfico estadístico *quantile-quantile (qq)*, que es un método gráfico para comparar dos distribuciones de probabilidad al trazar sus cuantiles uno contra otro; si las distribuciones son similares se esperaría que los puntos se acerquen a una recta diagonal. En este caso, comparamos los cuantiles de la distribución de los residuos contra una distribución normal con la media y desviación estándar que tienen los residuos.

  En la figura 4.2 vemos que la mayoría de los puntos trazados están sobre la recta diagonal, lo que nos permite asumir gráficamente que los residuos siguen una distribución normal. Analíticamente, para determinar si los residuos siguen una distribución normal se usarán los estadísticos definidos en la sección 2.1.3. Bajo la hipótesis nula de que los residuos siguen una distribución normal, por lo tanto un p-valor menor que 0,05 indicaría que se debe rechazar la hipótesis nula. Para el modelo de invierno los resultados de los estadísticos se presentan en el cuadro 4.3: la prueba de Jarque-Bera para los residuos presenta un p-valor mayor a 0,05 por lo cual no se puede rechazar el hecho de que los residuos siguen una distribución normal. Resultado similar se obtiene de la prueba de Anderson-Darling, cuyo estadístico $A = 0{,}707$ es menor que el valor crítico 0,752, presentado en el cuadro 2.1,

por lo cual no se rechaza la hipótesis nula de normalidad de los residuos; complementado con el hecho de que el p-valor de este estadístico es mayor que el nivel de significancia.

Figura 4.2: Gráfico *qq* de probabilidad normal de los residuos del modelo de invierno



Finalmente, el test de Shapiro-Wilk tampoco rechaza el supuesto de normalidad de los residuos.

Cuadro 4.3: Pruebas de normalidad del modelo de invierno

Test	Valor del estadístico	p-valor
Jarque-Bera	(χ) 3,0694	0,2155
Anderson-Darling	(A) 0,7075	0,0643
Shapiro-Wilk	(W) 0,99452	0,1376

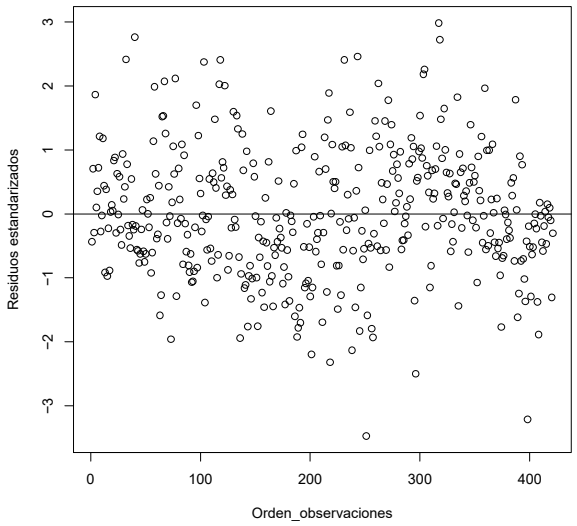
- **Autocorrelación:** La autocorrelación se define como la correlación entre miembros de obsrvaciones ordenadas en el tiempo para datos de series de timepo y en el espacio para datos de corte transversal. Matemáticamente se expresa como

$$cov(u_i, u_j|x_i, x_j) = E(u_i u_j) = 0 \qquad i \neq j. \tag{4.2}$$

Cuando los datos son muestras aleatorias de corte transversal se esperaría que no exista correlación entre una observación con otra, pero en caso de existir esta se

conoce como correlación espacial [12]. En el capítulo 1, en la sección 1.3, se presenta el procedimiento de recolección de datos realizado por el MAGAP, de manera que los datos recolectados son aleatorios y bajo esta premisa se espera que no exista correlación espacial. Una forma gráfica de verificar que en efecto los residuos no presentan atucorrelación es graficar los valores de los residuos respecto del número de orden de las observaciones. En la figura 4.3 se presenta los residuos estandarizados del modelo de regresión lineal múltiple para el ciclo de invierno, se puede observar que no hay patrón sistemático por lo que apoya el supuesto de no autocorrelación.

Figura 4.3: Gráfico de los residuos del modelo de invierno



- **Heterocedasticidad:** Para realizar el análisis estadístico de se usará las pruebas descritas en la sección 2.1.3. Se toma el p-valor tanto de la prueba de Breusch-Pagan como de la prueba de White. En el cuadro 4.4 se presentan los p-valores de cada test. Dado que los p-valores de los test de Breusch-Pagan y White no son menores que 0,05, el nivel de significancia elegido, podemos asumir que no existe evidencia estadística de heterocedasticidad.

Cuadro 4.4: Test de homocedasticidad del modelo de verano

Test	p-valor
Breusch-Pagan	0,1062
White	0,2650

Interpretación de los parámetros del modelo

Analizando los signos de los coeficientes, se puede aceptar que son congruentes con lo esperado. Hay que considerar que existen dos coeficientes con signo negativo, el de CI2 y el de VI1, que son variables por clúster de canton y variedad respectivamente. Se interpretan los coeficientes de las variables de la siguiente manera.

1. El uso del Kit del programa Plan Semilla de Alto Rendimiento por parte del agricultor aumenta el rendimiento en promedio en 0,05 Tm/ha respecto de los agricultores que no tuvieron acceso a este beneficio.

2. El aumento del número de mazorcas en diez metros lineales en una unidad, considerando el procedimiento determinado por el Manual del Investigador con el que se levanta la información, produce un aumento en el rendimiento en promedio en 0,49 Tm/ha, en comparación con los productores que registraron una unidad menos.

3. El uso de un quintal más de nitrógeno como fertilizante en el ciclo de cultivo aumenta el rendimiento en promedio en 0,08 Tm/ha.

4. Los agricultores que pertenecen a los cantones del cluster cantonal CI1 tienen en promedio un rendimiento de 0,05Tm/ha más que los agricultores de los demás clústeres cantonales.

5. Los agricultores del clúster cantonal CI2 tienen un rendimiento esperado de 0,07 Tm/ha menor al de los agricultores de los demás clústeres cantonales.

6. Los agricultores que usaron las variedades del clúster por variedad VI1 tienen un rendimiento esperado de 0,08 Tm/ha menor al de los agricultores que usaron otras vaiedades.

4.1.2. Modelo de RLM para el ciclo verano

Para el ciclo de verano se dispone de 642 encuestas a agricultores, de donde se obtienen cada una de las observaciones. Se tomaron los datos de verano correspondientes a los dos primeros ciclos para obtener el modelo de regresión y para el proceso de entrenamiento de las redes neuronales; esta muestra representa aproximadamente el 72 % de los registros de verano, en concreto, son 465 datos y los 177 datos restantes correspondientes serán usados para el test de los modelos.

Estimación de los parámetros del modelo

Considerando la descripción de la temperatura promedio del ciclo presentada en el cuadro 3.1 y la precipitación acumulada presentada en el cuadro 3.2, vemos que no superan los valores máximos recomendados respectivamente; en consecuencia, no es necesario incluir un polinomio cuadrático de estas variables en el modelo. La fertilización tampoco supera los valores máximos establecidos en la sección anterior, por lo tanto, no es necesario incluir su polinomio en el modelo.

Se determinaron las interacciones entre ser miembro de una asociación con las clusterizaciones por provincia y con el gasto en producción; los clústeres de variedad con la fertilización aplicada y con el número de mazorcas. El modelo obtenido mediante regresión por pasos a un nivel de significancia del 5 % presentado en el cuadro 4.5 muestra que la variable de interacción, notada var_ver_3_nm, entre el clúster VV3 con la variable N_M es significativa.

Cuadro 4.5: Modelo de RLM de verano

Coeficientes	Estimación	Error Std.	valor t	Pr($>\|t\|$)
N_M	0,35878	0,05085	7,056	6,40e-12 ***
CV2	0,10230	0,01509	6,779	3,77e-11 ***
T_P	0,16777	0,02471	6,789	3,52e-11 ***
G_P	0,11603	0,03651	3,178	0,00158 **
VV2	0,07601	0,01822	4,171	3,63e-05 ***
M_A	-0,03383	0,01383	-2,446	0,01482 *
VV3	0,04348	0,01887	2,304	0,02167 *
QQ_P	0,05953	0,02452	2,428	0,01557 *
var_ver_3_nm	0,14118	0,05417	2,606	0,00945 **
Observaciones	465			
R^2	0,9230			
R^2 Adjustado	0,9215			
Error Std. Residual	0,152(df = 456)			
Nota:	*p<0,05; **p<0,01, ***p<0,001			

Validación del modelo

- **Prueba RESET:** Para determinar la correcta especificación del modelo de verano, se realizó la prueba de error de especificación RESET. Considerando los p-valores del cuadro 4.6, no se rechaza la correcta especificación de la forma funcional del

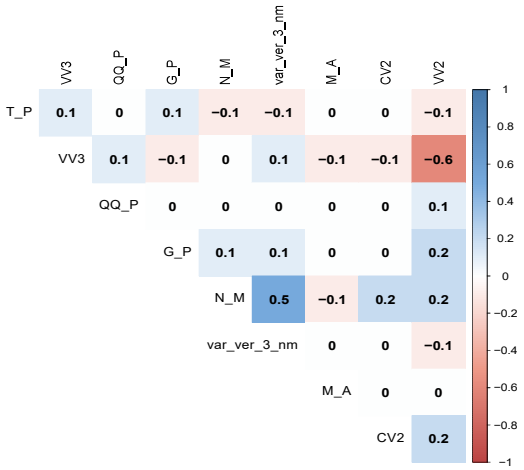
modelo.

Cuadro 4.6: Prueba RESET del modelo de verano

Potencia	P-valor
$\hat{y}^2$	0,3498
$\hat{y}^3$	0,3348
$\hat{y}^2, \hat{y}^3$	0,6253

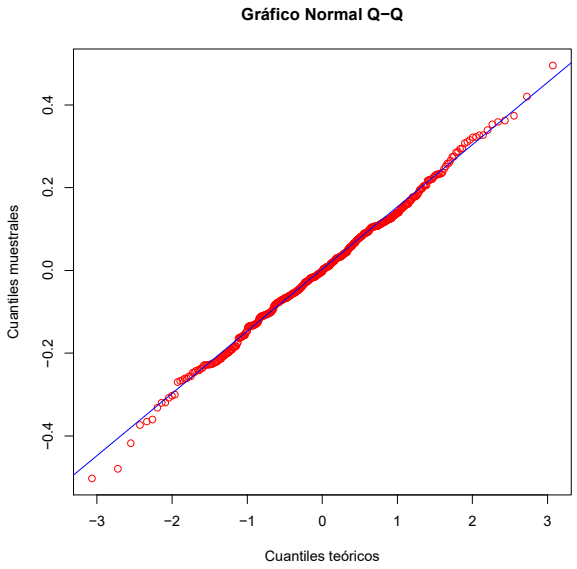
- **Multicolinealidad:** Podemos ver en la Figura 4.4 que no existe pareja de variables que estén altamente correlacionadas, dado que no existen valores superiores o cercanos a 0.9.

Figura 4.4: Matriz de correlación del modelo de verano



- **Normalidad:** El modelo tratado cumple con las pruebas de normalidad; como se puede observar en el gráfico quantile-quantile de la figura 4.5 los residuos siguen una distribución normal. Bajo la hipótesis nula de que los residuos siguen una distribución normal, se realizaron las pruebas de Jarque-Bera, Anderson-Darling y Shapiro-Wilk; considerando que las tres pruebas tienen p-valores mayores que 0,05, no se puede rechazar el hecho de que los residuos siguen una distribución normal.

Figura 4.5: Gráfico *qq* de probabilidad normal de los residuos del modelo de verano



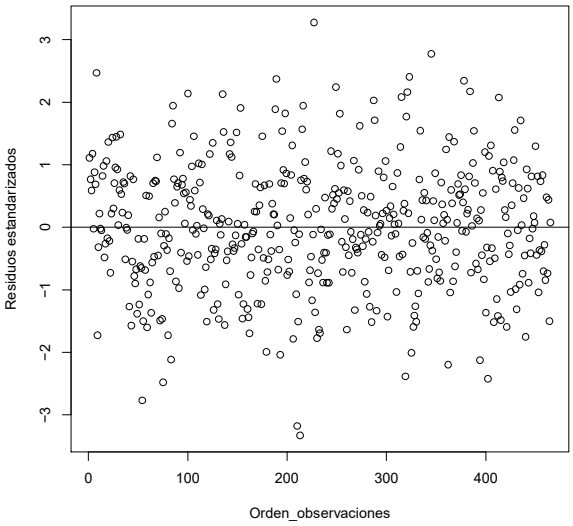
Cuadro 4.7: Pruebas de normalidad del modelo de verano

Test	Valor del estadístico	p-valor
Jarque-Bera	(χ) 1,431	0,4889
Anderson-Darling	(A) 0,362	0,4413
Shapiro-Wilk	(W) 0,997	0,6873

- **Autocorrelación:** En la sección 4.1.1 se explicó en qué consiste la autocorrelación de los residuos, ademas de especificar que para los datos del estudio una manera de apoyar el supuesto de no autocorrelación es mediente el gráfico de los residuos estandarizados respecto del orden de la muestra. De manera similar, para el modelo de RLM de verano se presenta en figura 4.6 residuos estandarizados, se puede observar que no hay patrón sistemático, hecho que apoya el supuesto de no autocorrelación[3].

[3]Notar que no se aplica la Prueba d de Durbin-Watson, dado que los datos corresponden a cortes transversales y no a una serie de tiempo, por tal no podemos tratar como correlación serial. Además, una de las condiciones para poder aplicar esta prueba no se cumple: modelo de regresión con el término del intercepto.

Figura 4.6: Gráfico de los residuos del modelo de verano



- **Heterocedasticidad:** El supuesto de homocedasticidad se cumple, tal como se observa en el cuadro 4.8 donde el p-valor del test de Breusch-Pagan y el test de White son mayores a 0,05, lo que implica no rechazar la hipótesis de homocedasticidad.

Cuadro 4.8: Test de homocedasticidad del modelo de verano

Test	p-valor
Breusch-Pagan	0,9059
White	0,1819

Interpretación de los parámetros del modelo

Analizando los signos de los coeficientes, se observa que tiene congruencia con lo esperado, a excepción del signo del coeficiente de ser miembro o no de una asociación, que si bien en un principio no resulta correcta, el análisis descriptivo de los datos y del fenómeno estudiado le dan sentido en cierta manera.

1. El aumento del número de mazorcas en diez metros lineales en una unidad, produce un aumento del rendimiento en promedio en 0,35 Tm/ha en comparación de los agricultores que registraron una unidad menos.

2. Los agricultores que pertenecen al clúster cantonal CV2 tienen en promedio un rendimiento de 0,10 Tm/ha más que los agricultores de los demás clústeres cantonales.

3. Un aumento en la temperatura en el ciclo de producción genera en promedio un aumento del rendimiento en 0,16 Tm/ha, considerando que los valores de temperatura registrados en este ciclo no superan el límite máximo sugerido por la CGSIN, en consecuencia, no se llega al punto en el que el aumento de temperatura deja de ser beneficioso para el cultivo.

4. Considerando que el gasto en producción se mide en miles de dólares por hectárea de cultivo, un aumento en el gasto de producción de mil dólares por hectárea genera un aumento del rendimiento en promedio del 0,11 Tm/ha.

5. Los agricultores que cultivan las variedades de maíz que pertenecen al clúster VV2 tienen un rendimiento promedio de 0,07 Tm/ha mayor al de los agricultores que cultivaron variedades diferentes a las de éste clúster.

6. Los agricultores que pertenecen a una asociación tienen un rendimiento promedio esperado de 0,03 Tm/ha menor que los agricultores que no pertenecen a una asociación. Si bien podría considerarse incorrecto el signo, si vemos la descripción de esta variable en la sección 3.1.1, vemos que la participación de asociados y no asociados es prácticamente del 50 %, pero destaca el hecho de que en verano los agricultores que siembran son los que poseen la infraestructura adecuada y mayor grado de tecnificación, lo que implica que los no asociados que siembran en verano tendrán un rendimiento igual o ligeramente mayor.

7. Los agricultores que cultivan las variedades de maíz que pertenecen al clúster por variedad VV3 tienen un rendimiento promedio de 0,04 Tm/ha mayor al de los agricultores que usan otras variedades y que están en la media del número de mazorcas, esto debido a la interección significativa de esta variable con el número de mazorcas.

8. El aumento en el uso de un quintal de fósforo por hectárea de cultivo genera un aumento en promedio de 0,059 Tm/ha.

4.2. Desarrollo de un modelo de redes neuronales

La selección de las variables es una etapa importante al momento de realizar el entrenamiento de una red neuronal artificial. Si bien podría considerarse que mientras más

entradas se agreguen mejor sería el desempeño de la red, es posible que su desempeño sea peor que con menos variables debido a que pueden ingresar variables que no aportan al modelo o bien están correlacionadas entre ellas [32],[38]. Por el contrario, se debe tener mucho cuidado en la selección y estracción de las variables a fin de conservar la información necesaria para el entrenamiento.
Cada investigador utiliza un método distinto para la selección de las variables de entrada de una red neuronal con el fin de evitar los problemas descritos; en el trabajo de investigación de Niazian *et al.* [32], en el cual estima el rendimiento de la semilla de Ajowan, las variables de entrada se seleccionaron mediante análisis de regresión por pasos. Hay que considerar que estas mismas variables son las seleccionadas para su modelo de RLM y ANN. Üreyen & Gürkan [47] en su trabajo de investigación sobre la estimación de propiedades de hilado mediante RLM y ANN toman como variables de entrada para las redes neuronales las ya determinadas por la RLM. Bilgili [2] considera todas las variables disponibles y realiza regresión múltiple por pasos para determinar el mejor modelo de regresión; y, propone tomar estas variables independientes seleccionadas en el modelo de RLM como entradas para el modelo de ANN.

Cuando no se opta por realizar análisis de regresión por pasos para determinar las variables de entrada de las redes neuronales Priddy & Keller [38] sugieren realizar análisis de componentes principales para lo cual los autores citan a Hotelling [19], análisis de discriminante lineal o mapeo de Fisher para lo cual los autores citan a Fisher [11], análisis de componentes independientes para lo cual los autores citan a Hyvärinen [18], entre otros dependiendo del problema a tratar. El estudio de cada uno de estos métodos es extenso y queda fuera del alcance de este trabajo.

Cuando se realizó la RLM para los ciclos de invierno y verano se tomaron las variables obtenidas por la regresión por pasos. En el presente estudio se tomarán las mismas variables como entradas para las redes neuronales.

4.2.1. Red neuronal para el ciclo de invierno

Para el ciclo de invierno se consideran como señales de entrada las variables determinadas por el modelo de RLM de invierno, presentadas en el cuadro 4.1: K_B, N_M, QQ_N, CI1, CI2, VI1 y vr_i1_qqn. La arquitectura que se usará es la red perceptrón multicapa (MLP) con una sola capa oculta [4] con n_1 nodos, así como se tiene en la figura 4.7.

[4]El teorema de aproximación universal establece que solo se necesita una capa oculta para aproximar cualquier función continua [34],[46].

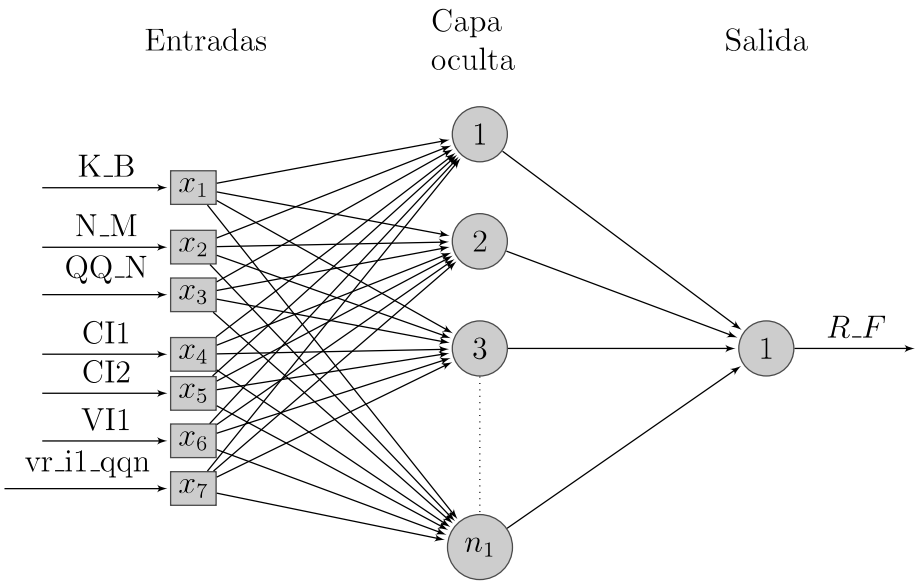
En esta etapa de determinación de la arquitectura óptima, se considera un intervalo del número de nodos que podría tener la única capa oculta. Según Nunes da Silva *et al* [34], para una red MLP con una sola capa oculta se considera la siguiente regla:

$$2 \cdot \sqrt{n} + n_2 \leq n_1 \leq 2 \cdot n + 1, \tag{4.3}$$

donde n es el número de entradas, n_2 es el número de neuronas de salida y n_1 es el número de neuronas en la capa oculta. En adelante se usará la notación $n - n_1 - n_2$ para nombrar a una arquitectura con n entradas, n_1 nodos en la capa oculta y n_2 nodos de salida. El máximo definido por este método evita la memorización de los datos de entrada (*sobreajuste*[5]), o la recolección insuficiente de información durante el proceso de entrenamiento (*desajuste*) [36],[37].

La elección de la función de activación es el siguiente paso a realizar; Niazian [32] recomienda realizar la normalización en el rango de la función de activación a usarse; dado que se normalizo en el rango $[0, 1]$, la función de activación a usar es la logística.

Figura 4.7: Red perceptrón multicapa para el ciclo de invierno



[5]El sobreajuste se produce cuando la red memoriza los datos de entrada, es decir, la red está muy alineada con un conjunto de datos en particular lo que provoca que para otros datos fuera de ese conjunto el ajuste sea deficiente [5], [15].

Algoritmo Backpropagation

En la inicialización del algoritmo backpropagation se debe elegir la tasa de aprendizaje η. Se comparó los estadísticos de bondad de ajuste considerando la arquitectura generada por las cotas inferior y superior de la ecuación 4.3, que son 6 y 15 respectivamente, así $\eta = 0{,}01$ demostró ser una tasa que aprendizaje alrededor de la cual las diferentes arquitecturas convergen. Los resultados de los estadísticos descritos en la sección 2.1.4 para cada arquitectura se presentan en el cuadro 4.9, tanto de la fase de entrenamiento como de la evaluación en el conjunto de test.

Cuadro 4.9: Arquitecturas para el ciclo invierno con el Algoritmo Backpropagation

Arquitectura	Entrenamiento			Test		
	r^2	$RMSE$	MAE	r^2	$RMSE$	MAE
7-6-1	0,4611	0,1245	0,0972	0,4547	0,1835	0,1503
7-7-1	0,4572	0,1250	0,0978	0,5041	**0,1640**	0.1347
7-8-1	0,4625	0,1243	0,0972	0,4757	0,1754	0,1407
7-9-1	0,4618	0,1244	0,0973	0,4531	0,1900	0,1556
7-10-1	0,4628	0,1243	0,0973	0,4470	0,1987	0,1633
7-11-1	0,4662	0,1239	0,0966	0,4935	0,1652	0,1346
7-12-1	0,4562	0,1251	0,0980	0,4800	0,1665	0,1347
7-13-11	0,4738	0,1230	0,0960	0,4706	0,1726	0,1408
7-14-1	**0,4810**	**0,1222**	**0,0948**	0,4094	0,1944	0,1587
7-15-1	0,4618	0,1244	0,0972	**0,5096**	0,1688	**0,1337**

Procedemos a analizar los resultados de los estadísticos del cuadro 4.9: considerando los estadísticos $r^2 = 0{,}4810$, $RMSE = 0{,}1222$ y $MAE = 0{,}0948$ se concluye que la mejor arquitectura de entrenamiento es la 7-14-1. Analizamos el comportamiento de las diferentes arquitecturas en el conjunto de test: la estructura 7-15-1 reporta el mejor error medio absoluto $MAE = 0{,}1337$, y el mayor coeficiente de correlación $r^2 = 0{,}5096$, por lo que es una de las principales arquitecturas a considerar; sin embargo, el mejor $RMSE$ se produce con la arquitectura 7-7-1, con un valor de 0,1640.

Algoritmo Rprop+

En el cuadro 4.10 se puede analizar que en la fase de entrenamiento la arquitectura 7-15-1 genera los mejores valores de dos estadísticos: $r^2 = 0,5482$, $RMSE = 0,1140$

y MAE = 0,0878. Adicionalmente, en la fase de test se observa que la arquitectura 7-14-1 presenta los mejores los mejores estadísticos $r^2 = 0,4818$, $RMSE = 0,1701$ y MAE = 0,1378. Con éste algoritmo se utilizará la arquitectura 7-14-1.

Cuadro 4.10: Arquitecturas para el ciclo invierno con el Algoritmo Rprop+

Arquitectura	Entrenamiento			Test		
	r^2	$RMSE$	MAE	r^2	$RMSE$	MAE
7-6-1	0,5017	0,1198	0,0932	0,3807	0,1819	0,1490
7-7-1	0,4892	0,1212	0,0948	0,3953	0,1834	0,1496
7-8-1	0,5465	0,1142	0,0885	0,3103	0,1925	0,1499
7-9-1	0,5428	0,1147	0,0880	0,3952	0,1797	0,1464
7-10-1	0,5379	0,1153	0,0889	0,4190	0,1786	0,1433
7-11-1	0,5232	0,1171	0,0906	0,3366	0,1874	0,1531
7-12-1	0,4895	0,1212	0,0943	0,3903	0,1872	0,1512
7-13-1	0,4959	0,1204	0,0936	0,4045	0,1787	0,1453
7-14-1	0,4897	0,1212	0,0942	**0,4818**	**0,1701**	**0,1378**
7-15-1	**0,5482**	**0,1140**	**0,0878**	0,3445	0,2037	0,1562

Algoritmo Rprop-

A partir de los resultados presentados en el cuadro 4.11, se determina que la mejor arquitectura de la fase de entrenamiento es la 7-13-1 con r^2 = 0,5545, $RMSE$ = 0,1132 y MAE = 0,0866 Luego, realizando las estimaciones con el conjunto de test, se constata que la arquitectura 7-12-1 con r^2 = 0,5014, $RMSE$ = 0,1704 y MAE = 0,1372 es la que mejor ajusta los datos. Por sus mejores resultados se considera la arquitectura 7-12-1 a realizarce con el algoritmo Rprop-.

Cuadro 4.11: Arquitecturas para el ciclo invierno con el Algoritmo Rprop-

Arquitectura	Entrenamiento			Test		
	r^2	$RMSE$	MAE	r^2	$RMSE$	MAE
7-6-1	0,4907	0,1211	0,0938	0,4269	0,1742	0,1405
7-7-1	0,4935	0,1207	0,0941	0,3674	0,1826	0,1468
7-8-1	0,5051	0,1193	0,0926	0,4198	0,1861	0,1464
7-9-1	0,5231	0,1171	0,0907	0,4131	0,2080	0,1698
7-10-1	0,5240	0,1170	0,0906	0,4049	0,1800	0,1477
7-11-1	0,5014	0,1198	0,0936	0,2832	0,1946	0,1602
7-12-1	0,4922	0,1209	0,0940	**0,5014**	**0,1704**	**0,1372**
7-13-1	**0,5545**	**0.1132**	**0,0866**	0,3012	0,1994	0,1605
7-14-1	0,5094	0,1188	0,0917	0,4667	0,1912	0,1514
7-15-1	0,5404	0,1150	0,0884	0,3989	0,2070	0,1597

Cuadro 4.12: Mejores arquitecturas para el ciclo de invierno

Algoritmo	Arquitectura	Entrenamiento			Test		
		r^2	$RMSE$	MAE	r^2	$RMSE$	MAE
backprop	**7-15-1**	0,4618	0,1244	0,0972	0,5096	0,1688	0,1337
Rprop +	7-14-1	0,4897	0,1212	0,0942	0,4818	0,1701	0,1378
Rprop -	**7-12-1**	0,4922	0,1209	0,0940	0,5014	0,1704	0,1372

Las arquitecturas que mejores predicciones realizaron con los tres algoritmo usados se presenta el en cuadro 4.12. Considerando estas tres arquitecturas, podemos ver que la 7-15-1 con el algoritmo backprop genera mejores resultados en la fase de test, sin embargo, en la fase de entrenamiento es la 7-12-1 con el algoritmo Rprop - la que genera mejores resultados. Si comparamos éstas arquitecturas, en la fase de test la diferencia porcentual en los estadísticos es de 0,82 % en r^2, 0,16 % en $RMSE$ y 0,35 % en MAE. Por otro lado, en la fase de entrenamiento la diferencia es de 3,04 % en r^2, 0,35 % en $RMSE$ y 0,32 % en MAE. Se concluye que la arquitectura 7-12-1 es ligeramente inferior en la fase de test, y un rendimiento mejor en la fase de entrenamiento en comparación con la arquitectura 7-15-1. Se considera ambas arquitecturas y toda la base del ciclo invierno; en el cuadro 4.13 se reporta los resultados obtenidos. Los resultados de los estadísticos determinan que la mejor arquitectura es la 7-12-1 con el algoritmo Rprop -.

Cuadro 4.13: Mejores arquitecturas para los datos de todo el ciclo de invierno

Algoritmo	Arquitectura	Todo el ciclo		
		r^2	$RMSE$	MAE
backprop	7-15-1	0,4400	0,1575	0,1453
Rprop -	7-12-1	0,4691	0,1533	0,1405

Resultados

Comparando los modelos mediante los estadísticos de bondad de ajuste, se puede observar en el cuadro 4.14 que tanto en la fase de entrenamiento como en la fase de test el modelo de redes neuronales presenta mejores estadísticos de bondad de ajuste. Para la fase de entrenamiento el modelo de ANN presenta una mejoría del 5,29 % del porcentaje de correlación, una disminución del error medio cuadrático del 0,61 % y disminución del error absoluto del 0,56 %. En la fase de test el resultado es mejor en los estadístios de error, con una disminución del 2,28 % en el error medio cuadrático y una disminución del 2,29 % en el error absoluto; el porcentade de correlación mejoró en un 4,49 %.

Cuadro 4.14: Comparación de los modelos ANN y RLM para el ciclo invierno

Modelo	Arquitectura	Entrenamiento			Test		
		r^2	$RMSE$	MAE	r^2	$RMSE$	MAE
ANN	7-12-1	0,4922	0,1209	0,0940	0,5014	0,1704	0,1372
RLM	-	0,4393	0,1270	0,0996	0,4565	0,1932	0,1601

Cuadro 4.15: Matriz de pesos del modelo de invierno

j	W_{j1}	W_{j2}	W_{j3}	W_{j4}	W_{j5}	W_{j6}	W_{j7}	θ_j
1	0,95272	-2,92335	-0,38437	1,75034	3,66114	-1,81688	0,75208	-0,37370
2	-0,85667	-1,18098	0,41918	-0,52787	-1,92749	1,76674	0,19712	0,80974
3	-0,25353	-2,01895	-1,41251	1,79790	-1,64974	-0,93168	-0,69173	1,71075
4	3,37078	1,76117	-7,89958	5,72891	-1,14228	-3,11251	-0,13019	0,62304
5	0,89951	0,52840	-1,01882	-12,13515	-0,88035	2,55010	-1,93959	-0,60503
6	-0,75788	-0,65383	0,01094	0,60056	1,39087	-0,35762	1,29474	-0,80561
7	0,77686	0,40882	-2,35584	4,54362	-0,14432	-4,73161	-0,73285	0,00028
8	3,01776	4,04564	-1,50198	0,14313	2,57386	0,48842	-1,16497	-2,61990
9	35,70424	3,06606	-0,93161	32,43865	35,69496	-1,12209	-1,58234	0,09322
10	0,11049	1,49386	-0,46897	-0,52136	0,43354	34,36889	-1,97730	1,93485
11	-7,54716	-8,59568	17,62003	1,06112	-1,37120	0,15213	0,68309	0,28654
12	1,41530	0,56755	0,26447	-0,75976	-1,15643	-2,33544	-1,31823	0,56558

Para determinar la expresión matemática del modelo ANN para el ciclo invierno, se determina primero la matriz de pesos $W^{(1)}_{12,7}$, cuyos valores son los pesos que conectan la neurona j de la capa oculta a la neurona i de la capa de entrada, donde $j = 1, \ldots, 12$ e $i = 1, \ldots, 7$. Notar que esta matriz incluye los umbrales de activación θ_j. Los valores de la matriz $W^{(1)}_{12,7}$ se pueden observar en el cuadro 4.15 junto a los umbrales de activación.

La ecuación final del modelo de red neuronal artificial para el ciclo de invierno se establece en la siguiente ecuación:

$$\begin{aligned} R_F = &- 1{,}66726g(u_1) - 0{,}61865g(u_2) - 1{,}60820g(u_3) + 1{,}02356g(u_4) \\ &- 2{,}48322g(u_5) - 1{,}27393g(u_6) + 0{,}69786g(u_7) + 1{,}34848g(u_8) \\ &- 0{,}84864g(u_9) + 0{,}74623g(u_{10}) + 0{,}88457g(u_{11}) \\ &- 0{,}57786g(u_{12}) + 0{,}97287 \end{aligned} \tag{4.4}$$

Donde $g(u_j)$, para $j = 1, \ldots, 12$ se define como

$$g(u_j) = \frac{1}{1 + e^{-u_j}}. \tag{4.5}$$

Los valores de u_i para las entradas de la ecuación 4.5 se calculan usando la matriz de pesos del cuadro 4.15 y los valores de las entradas de la red, de la siguiente manera:

$$\begin{aligned} u_j = &K_B \cdot W_{j1} + N_M \cdot W_{j2} + QQ_N \cdot W_{j3} + CI1 \cdot W_{j4} \\ &+ CI2 \cdot W_{j5} + VI1 \cdot W_{j6} + vr_i1_qqn \cdot W_{j7} + \theta_j \end{aligned} \tag{4.6}$$

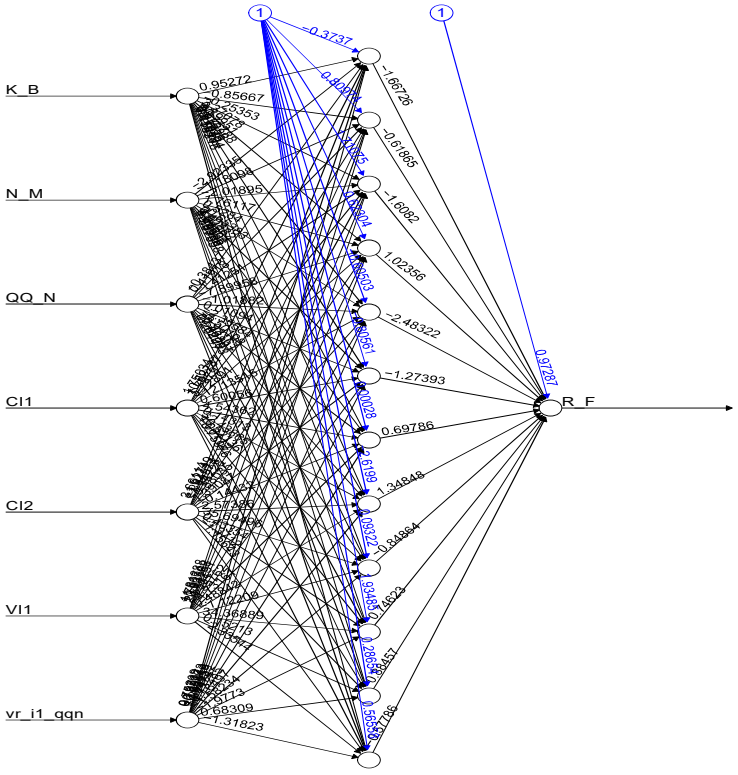
En la figura 4.8 se puede observa las conecciones neuronales junto con los pesos de la red neuronal de arquitectura 7-12-1 para el ciclo de invierno.

4.2.2. Red neuronal para el ciclo de verano

Para determinar la red neuronal del rendimiento de maíz en el ciclo de verano se usará como entradas las variables determinadas mediante la RLM realizada en la sección anterior. Es así que las variables N_M, CV2, T_P, G_P, VV2, M_A, VV3, QQ_P y var_ver_3_nm representarán las entradas de la red; la arquitectura a entrenar es la que se presenta en la figura 4.9.

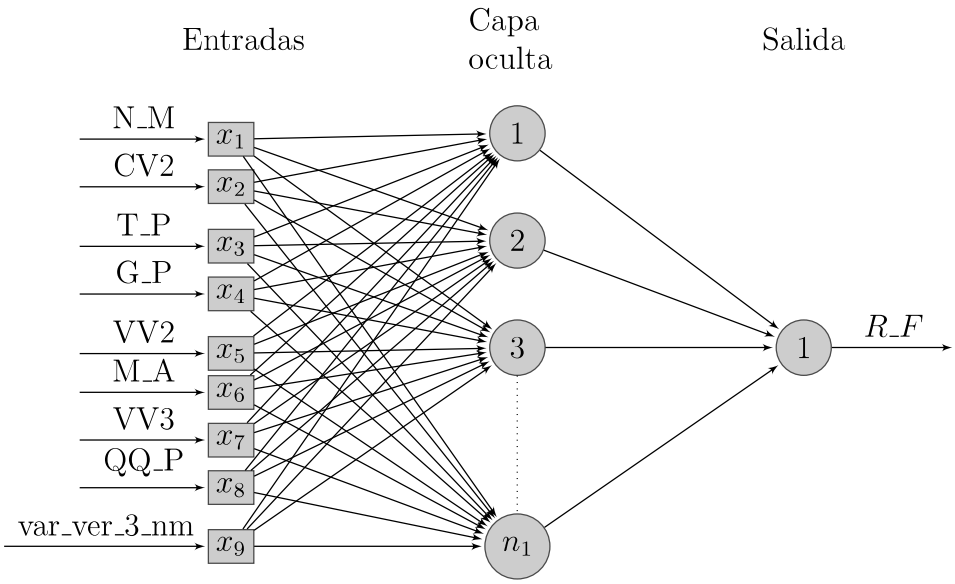
Además, se determina la cota inferior y superior de la ecuación 4.3, las cuales son 7 y

Figura 4.8: Modelo de red neuronal para el ciclo de invierno



19 respectivamente. La función de activación a usar será la logística.

Figura 4.9: Red perceptrón multicapa para el ciclo de verano



Algoritmo Backpropagation

Para el entrenamiento de la red neuronal la tasa de aprendizaje $\eta = 0{,}01$ demostró ser consistente con la cual las diferentes arquitecturas convergen, a excepción de las arquitecturas con más de diez nodos en la capa oculta y ajustada mediante el algoritmo backpropagation. Los resultados de los estadísticos para cada arquitectura se presentan en el cuadro 4.16, tanto de la fase de entrenamiento como de la evaluación en los conjuntos de test. A partir de estos resultados, la arquitectura 9-9-1 es la que presenta los mejores resultados en la fase de entrenamiento con valores de $r^2 = 0{,}5577$, $RMSE = 0{,}1277$ y $MAE = 0{,}1000$. En la fase de test, la estructura 9-9-1 presenta el mejor $r^2 = 0{,}1765$, con los estadísticos $RMSE = 0{,}2028$ y $MAE = 0{,}1638$.

Cuadro 4.16: Arquitecturas para el ciclo verano con el Algoritmo Backpropagation

Arquitectura	Entrenamiento			Test		
	r^2	$RMSE$	MAE	r^2	$RMSE$	MAE
9-7-1	0,5130	0,1397	0,1108	0,1525	**0,2022**	**0,1593**
9-8-1	0,5393	0,1337	0,1056	0,1589	0,2053	0,1645
9-9-1	**0,5577**	**0,1277**	**0,1000**	**0,1765**	0,2028	0,1638

Algoritmo Rprop+

En la fase de entrenamiento de este algoritmo tenemos que la arquitectura 9-19-1 genera los mejores resultados: $r^2 = 0{,}7817$, $RMSE = 0{,}0898$ y $MAE = 0{,}0669$. A considerar, en la fase de test la arquitectura con mejor $r^2 = 0,1412$ es la 9-9-1; la 9-8-1 determina la major $RMSE = 0{,}2138$ y la 9-9-1 presenta la menor $MAE = 0{,}1681$, así como se puedo observar en el cuadro 4.17. Se toma la arquitectura 9-9-1 como la mejor con este algoritmo.

Cuadro 4.17: Arquitecturas para el ciclo verano con el Algoritmo Rprop+

Arquitectura	Entrenamiento			Test		
	r^2	$RMSE$	MAE	r^2	$RMSE$	MAE
9-7-1	0,5802	0,1244	0,0999	0,1136	0,2170	0,1737
9-8-1	0,5939	0,1224	0,0946	0,1204	**0,2138**	0,1722
9-9-1	0,6099	0,1200	0,0933	**0,1412**	0,2141	**0,1681**
9-10-1	0,6100	0,1199	0,0914	0,1251	0,2180	0,1779
9-11-1	0,6764	0,1092	0,0830	0,1361	0,2303	0,1878
9-12-1	0,6599	0,1120	0,0865	0,0958	0,2353	0,1934
9-13-1	0,7031	0,1046	0,0810	0,0934	0,2244	0,1775
9-14-1	0,6747	0,1095	0,0851	0,1074	0,2220	0,1787
9-15-1	0,7257	0,1006	0,0776	0,1326	0,2271	0,1839
9-16-1	0,7263	0,1005	0,0783	0,0908	0,2660	0,2214
9-17-1	0,7289	0,1000	0,0749	0,1241	0,2359	0,1919
9-18-1	0,7293	0,0999	0,0764	0,1024	0,2397	0,1944
9-19-1	**0,7817**	**0,0898**	**0,0669**	0,0586	0,2625	0,2178

Algoritmo Rprop-

En la fase de entrenamiento con este algoritmo, cuyos resultados se presentan en el cuadro 4.18, la arquitectura 9-19-1 genera los mejores resultados: $r^2 = 0{,}7683$, $RMSE = 0{,}0924$ y $MAE = 0{,}0687$. Al realizar el test se observa que la mejor arquitectura considerando r^2 es la 9-8-1 con un $r^2 = 0{,}1803$, $RMSE = 0{,}2014$ y $MAE = 0{,}1602$.

Finalmente, consideramos las mejores arquitecturas entrenadas con los diferentes algoritmos, dichas arquitecturas se presentan en el cuadro 4.19. Se puede observar que el algoritmo Rprop - con la arquitectura 9-8-1 genera los mejores estadísticos de bondad de ajuste, en todo caso será éste el modelo de red neuronal para el ciclo de verano.

Cuadro 4.18: Arquitecturas para el ciclo verano con el Algoritmo Rprop-

Arquitectura	Entrenamiento			Test		
	r^2	$RMSE$	MAE	r^2	$RMSE$	MAE
9-7-1	0,5519	0,1286	0,1010	0,1474	0,2076	0,1666
9-8-1	0,5636	0,1269	0,0981	**0,1803**	**0.2014**	**0.1602**
9-9-1	0,6090	0,1201	0,0944	0,1163	0,2229	0,1739
9-10-1	0,6123	0,1196	0,0928	0,1295	0,2118	0,1717
9-11-1	0,6327	0,1164	0,0898	0,1216	0,2263	0,1860
9-12-1	0,6614	0,1117	0,0862	0,1337	0,2189	0,1788
9-13-1	0,6460	0,1143	0,0877	0,1225	0,2292	0,1836
9-14-1	0,7035	0,1046	0,0796	0,1251	0,2344	0,1910
9-15-1	0,7193	0,1017	0,0788	0,0439	0,2440	0,1950
9-16-1	0,7233	0,1010	0,0774	0,0255	0,2677	0,2209
9-17-1	0,7423	0,0975	0,0749	0,0195	0,2845	0,2308
9-18-1	0,7518	0,0957	0,0718	0,0905	0,2547	0,2067
9-19-1	**0,7683**	**0,0924**	**0,0687**	0,0743	0,2690	0,2193

Cuadro 4.19: Mejores arquitecturas para el ciclo de verano

Algoritmo	Arquitectura	Entrenamiento			Test		
		r^2	$RMSE$	MAE	r^2	$RMSE$	MAE
backprop	9-9-1	0,5577	0,1277	0,1000	0,1765	0,2028	0,1638
Rprop +	9-9-1	0,6099	0,1200	0,0933	0,1412	0,2141	0,1681
Rprop -	9-8-1	0,5636	0,1269	0,0981	0,1803	0,2014	0,1602

Resultados

Realizando una comparación los modelos mediante los estadísticos de bondad de ajuste, se puede observar en el cuadro 4.20 que tanto en la fase de entrenamiento el modelo de redes neuronales presenta mejores resultados, as así que para el estadístico r^2 presenta una mejoría porcentual del 18,01 %, una disminución del error medio cuadrático en 2,4 % y disminución del error absoluto en 2,05 %. En la fase de test el porcentaje de correlación mejoró en un 1,5 %; en los estadístios de error por el contrario se nota un ligero aumento con un 0,19 % en el error medio cuadrático y un 0,28 % en el error absoluto.

La matriz de pesos denotada $W^{(1)}_{8,9}$, contiene los pesos que conectan la neurona j de

Cuadro 4.20: Comparación de los modelos ANN y RLM para el ciclo verano

Modelo	Arquitectura	Entrenamiento			Test		
		r^2	$RMSE$	MAE	r^2	$RMSE$	MAE
ANN	9-8-1	0,5636	0,1269	0,0981	0,1803	0,2014	0,1602
RLM	-	0,3835	0,1509	0,1186	0,1653	0,1995	0,1574

la capa oculta a la neurona i de la capa de entrada, donde $j = 1, \ldots, 8$ e $i = 1, \ldots, 9$ y además contiene los umbrales de activación θ_j. Los valores de la matriz $W_{8,9}^{(1)}$ se pueden observar en el cuadro 4.21.

Cuadro 4.21: Matriz de pesos del modelo de verano

j	W_{j1}	W_{j2}	W_{j3}	W_{j4}	W_{j5}	W_{j6}	W_{j7}	W_{j8}	W_{j9}	θ_j
1	0,30521	0,52995	-0,51954	3,31994	-0,59765	-245,77704	-2,66880	0,57883	-2,27117	-1,86849
2	0,86971	-7,03863	9,52907	-10,05292	3,28867	6,18036	43,58670	4,77006	1,88986	2,51184
3	-1,13362	-0,58271	-1,62872	-0,86308	0,00071	0,77030	-0,55722	0,73734	-0,91100	2,13489
4	-1,20199	5,11833	-15,24424	1,90278	0,62107	0,47205	4,67920	-6,52350	2,04576	-5,99076
5	7,22425	-1,58571	-4,58491	-0,30910	2,83143	-0,94924	0,34010	-1,06743	-2,48991	-0,33936
6	-0,26922	1,42038	-5,79344	-2,78552	0,81866	-0,73593	2,15964	-2,00981	1,21819	-0,02738
7	-9,65855	2,65919	-1,11713	0,66021	-3,56342	4,10161	-0,78216	4,94074	-0,00919	3,75598
8	0,35527	-2,44782	4,37191	-8,56374	1,66923	-2,44304	-1,31357	39,96715	3,43304	5,69678

La ecuación final del modelo de red neuronal artificial para el ciclo de verano se establece de la siguiente manera:

$$\begin{aligned} R_F = &- 2{,}23244g(u_1) - 1{,}19707g(u_2) - 3{,}28329g(u_3) + 2{,}87014g(u_4) \\ &+ 1{,}71584g(u_5) - 2{,}53031g(u_6) + 1{,}11255g(u_7) + 2{,}06472g(u_8) - 0{,}31262 \end{aligned}$$

Donde $g(u_j)$, para $j = 1, \ldots, 8$ se define como

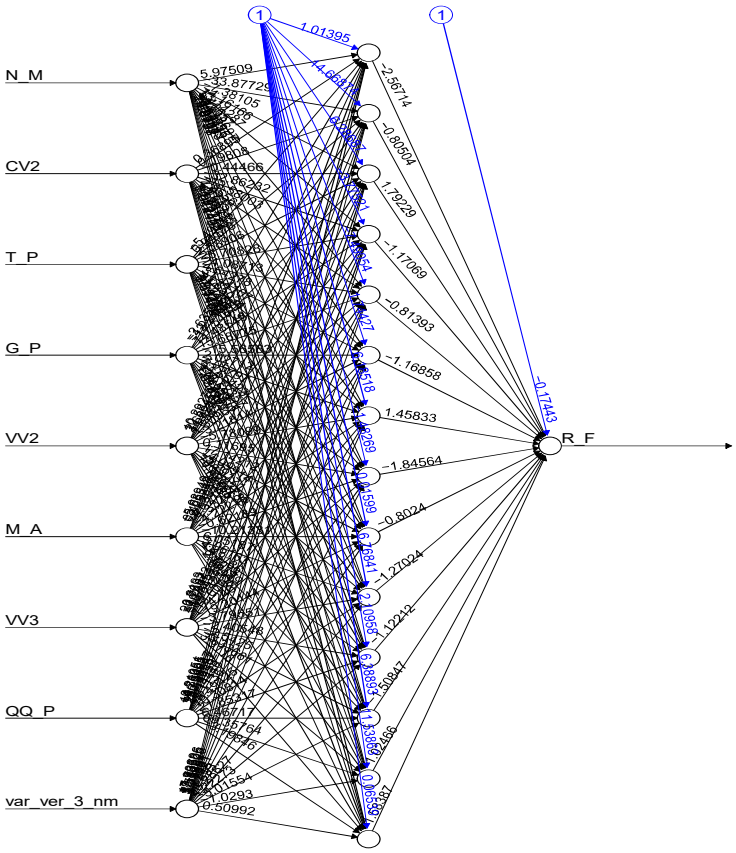
$$g(u_j) = \frac{1}{1 + e^{-u_j}}. \tag{4.7}$$

Los valores de u_i se calculan usando la matriz de pesos del cuadro 4.21 junto con los valores de las entradas de la red, de la siguiente manera:

$$\begin{aligned} u_j = &N_M \cdot W_{j1} + CV2 \cdot W_{j2} + T_P \cdot W_{j3} + G_P \cdot W_{j4} \\ &+ VV2 \cdot W_{j5} + M_A \cdot W_{j6} + VV3 \cdot W_{j7} + QQ_P \cdot W_{j8} + var_ver_3_nm \cdot W_{j9} + \theta_j \end{aligned} \tag{4.8}$$

En la figura 4.10 se puede observa las conecciones neuronales junto con los pesos de la red neuronal de arquitectura 7-12-1 para el ciclo de invierno.

Figura 4.10: Red neuronal ajustada para el ciclo de verano



Capítulo 5

Conclusiones y recomendaciones

- En el proceso de cultivo de maíz intervienen varios factores, entre ellos los climáticos, la ubicación geográfica, la cantidad de fertilizante usado, la variedad utilizada, la inversión realizada entre otros. Usualmente un modelo de regresión lineal múltiple es usado para la estimación del rendimiento, considerando que se cumplen los supuestos necesarios para la validación de los resultados; sin embargo, este modelo no considera las relaciones no lineales que puede existir entre los factores de cultivo y el rendimiento. Un modelo de redes neuronales artificiales se ajusta a relaciones no lineales por lo cual es una mejor opción al momento de modelar el rendimiento de maíz.

- La diferencia estructural en el comportamiento del rendimiento de maíz en cada ciclo de cultivo llevó a tratar dos modelos, uno para cada ciclo, tanto de regresión lineal múltiple como de redes neuronales artificiales. A diferencia de los modelos propuestos por la CGSIN [30] en 2014, este trabajo considera variables adicionales de clusterización por provincias, cantones y variedad; adicional a esto, se consideraron interacciones entre la variedad contra la fertilización y el número de mazorcas en el ciclo de invierno, y para el ciclo de verano se consideró la interacción entre ser miembro de una asociación contra los clústeres de provincia y el gasto en producción; y la interacción entre la variedad contra la fertilización y el número de mazorcas. Los modelos así determinados cumplen los supuestos del modelo clásico de regresión por lo que sus resultados son válidos.

- El modelo de RLM del ciclo de invierno determinó que las variables de fertilización, posición geográfica y variedad tienen una mayor incidencia en el rendimiento de maíz por hectárea. Además de estas variables, otras que resultaron ser influyentes

son el número de mazorcas registradas en 10 metros lineales y la interacción entre la variedad sembrada y la cantidad de nitrógeno usado.

- El rendimiento presenta una relación en forma de U invertida con las variables de temperatura, precipitación y fertilización, sin embargo , del análisis descriptivo se determinó que no son necesarios incluir en los modelos de RLM los términos cuadráticos de dichas variables; así mismo, mediante la prueba RESET se determina que los términos cuadráticos de éstas variables no son estadísticamente significativos.

- En el ciclo de invierno se determinó que, *ceteris paribus*, el fertilizante que más contribuye al aumento del rendimiento del maíz es el nitrógeno; considerando que el valor máximo de quintales de nitrógeno usado es de 4,14qq/ha y que lo recomentado por la INIAP es usar entre 5 a 7 qq/ha, aún se podría mejorar el rendimiento con el aumento de este fertilizante. Además, su influencia también depende de la variedad usada.

- Para el ciclo de invierno el uso del Kit del Progama Plan Semilla de Alto Rendimiento es un factor influyente en el rendimiento del maíz, esto también determina que la ayuda gubernamental es importante considerando que en este ciclo hay mayor participación en la siembra por parte de los pequeños y medianos productores.

- Para el ciclo de verano se determina que, *ceteris paribus*, la temperatura promedio del ciclo es de mayor influencia en el rendimiento de maíz, reflejando que a un aumento de temperatura mejor será el rendimiento, pero hay considerar que en este ciclo la temperatura máxima y mínima bordean los 25° que es lo óptimo recomendado y está lejos de los 35° a partir de lo cual deja de ser beneficioso.

- El gasto en la producción es otro factor importante en el rendimiento del maíz, y como es de esperarse, el aumento en el gasto para el cultivo permite un aumento en el rendimiento.

- En el ciclo de verano se determinó que, *ceteris paribus*, potasio es el fertilizante que más contribuye al aumento del rendimiento de maíz. Considerando que una cantidad adecuada de postasio ayuda a las plantas a soportar sequías, plagas y enfermedades es de esperarse que un mayor uso de este fertilizante mejore el rendimiento; no obstante, la cantidad máxima usada para este ciclo es de 1,6 qq/ha, que está fuera de lo recomendado que es de 2 a 3 qq/ha, por lo que aún se puede mejorar el rendimiento aumentando el uso al intervalo recomendado.

- El modelo de verano refleja que, *ceteris paribus*, los agricultores que son miembros de una asociación tienen un rendimiento ligeramente menor de los demás. Uno de los factores que explican este comportamiento es el hecho de que en verano los agricultores que siembran son los que poseen la infraestructura adecuada y un mayor grado de tecnificación, sin embrago, sería importante investigar otros factores no analizados, como el grado de conocimiento agrícola del agricultor.

- Una de las desventajas de las redes neuronales es que no permite realizar un análisis *ceteris paribus* entre las variables de entrada de la red neuronal.

- Para el proceso de entrenamiento de las redes neuronales tanto de invierno como de verano, la tasa de aprendizaje $\eta = 0{,}01$ demostró ser consistente para la convergencia del algoritmo Backpropagation y sus versiones optimizadas Rprop + y Rprop - en las diferentes arquitecturas elegidas, considerando el número mínimo y máximo de nodos en la capa oculta.

- Para el ciclo de invierno la arquitectura 7-12-1 ejecutada con el algoritmo Rprop- es la que reporta valores ligeramente mejores para de los estadísticos de bondad de ajuste descritos en la sección 2.1.4, tanto en la fase de entrenamiento como en la fase de test en comparación con el modelo de RLM. Los resultados numéricos se presentan en la cuadro 4.14.

- Para la elección del mejor modelo de red neuronal para el ciclo de invierno fue necesario realizar el entrenamiento del modelo con todos los datos del ciclo, esto debido a que las arquitecturas 7-15-1 con el algoritmo backprop y la 7-12-1 con el algoritmo Rprop- presentan valores ligeramente distintos en la fase de test y no fue posible emitir un criterio de cuál es mejor.

- En el ciclo de verano la arquitectura 9-8-1 ejecutada con el algoritmo Rprop- es la reporta mejores valores de los estadísticos de bondad de ajuste en la fase de entrenamiento en comparación con los reportados por el modelo de RLM de este ciclo; en la fase de test solo el estadístico de correlación presenta una mejoría en el modelo de redes neuronales, los estadísticos de error presentaron un legiro incremento. Los resultados numéricos se los puede observar en el cuadro 4.20.

- La ubicación geográfica es influyente en el rendimiento de maíz, dado que las variables de clusterización por provincia o cantón son significativas en los modelos de RLM del ciclo de invierno y verano; y, considerando que una de las principales ca-

racterísticas geográficas en los cultivo es el tipo de suelo, sería recomendable incluir el tipo de suelo al momento de levantar los datos.

- Considerando que las entradas de las redes neuronales fueron tomadas a partir de las variables determinadas por la RLM, sería recomendable usar otras técnicas de selección de variables, como las descritas en la sección 4.2, a fin de determinar otras variables influyentes en el rendimiento de maíz.

Capítulo 6

Anexos

ANEXO 1

Ministerio de Agricultura, Ganadería, Acuacultura y Pesca

Observación No.: ----------

Ubicación de la finca o hacienda

Nombre Investigador: ---------- Fecha Investigación: ----------
Provincia: ---------- Cantón: ----------
Parroquia: ---------- Sitio o recinto: ----------

Operativo para comprobación y ajuste de rendimientos de maíz

VERANO 2014

Coordenadas Geográficas: X: ---------- Y: ---------- Z: ----------

Fecha de Investigación

Sección 1. Datos del productor

1.1. ¿Cuál es el nombre completo del productor? ---------- *Primer Nombre* ---------- *Segundo Nombre* ---------- *Primer Apellido* ---------- *Segundo Apellido*

1.2. ¿Cuál es el número de CC del productor? *Número de Cédula de Ciudadanía*

1.3. Números telefónicos de contacto. ---------- *Número convencional* ---------- *Número celular* ---------- *Otro número*

1.4. ¿Recibió alguna capacitación relacionada con la produccion del cultivo durante el último año? *1. Si:* *2.No:*

1.5. ¿Es miembro de alguna asociación? *1. Si:* *2.No:*

Si la respuesta anterior es Si, responda:
1.5.1 ¿Ha recibido algún beneficio para mejorar su producción durante el último año? *1. Si:* *2.No:*

Sección 2. Datos de la finca o hacienda

2.1. ¿Cuál es la superficie total de la finca o hacienda? *ha*

2.2. ¿Qué superficie dedica actualmente al cultivo de maíz? *ha*

2.3. ¿Cuál fue la fecha de siembra? *dia/mes/año*

2.4. ¿Cuál será la fecha de cosecha aproximada? *dia/mes/año*

2.5. ¿El cultivo es:? *1. Solo:* *2. Asociado:*

Si la respuesta anterior es asociado, responda:
2.5.1 ¿Con que otro producto siembra?

Sección 3. Encuesta respecto al producto

Siembra

3.1. ¿Qué material vegetativo utilizó para la siembra? *1. Semilla:* *2. Plántula:*

3.2. ¿De dónde obtuvo su semilla/plántula? *1. Reciclada Propia:* *2. Reciclada Otros:* *3. Comprada en casa comercial o semilleristas:*

Si la respuesta anterior fue semilla, responda:
3.2.1 ¿Qué cantidad de semilla utilizó por hectárea? *Cantidad:* *Presentación:* *Peso/Unidades:*

3.3. ¿Qué variedad o hibrido utilizó?

3.4. ¿Qué método de siembra utilizó? *1. Al voleo:* *2. Chorro continuo:* *3. Distanciamiento:*

3.5. ¿Utilizó el Kit del Plan Semilla en el cultivo del que se tomará la muestra? *1. Si:* *2.No:*

Crecimiento, Floración y Fructificación

3.6. ¿Cuántos quintales de fertilizante utilizó en el cultivo?

Fertilizante	*1.Urea*	*2. MOP*	*3. DAP*	*4. Mezclas*		
				4.1 Inicio	*4.2 Crecimiento*	*4.3.Llenado*
				Fosforo	*Nitrogeno*	*Potasio*
Quintales/ha.						

3.7. ¿Cuál de las siguientes labores se realizó de manera mecánica? *1.Preparación del suelo:* *2. Siembra:* *3. Fertilización:* *4. Control Malezas y enfermedades:* *5. Cosecha:*

3.8. ¿A que distancia aproximada se encuentra la fuente de agua más cercana? *km*

3.9. ¿Accede a la fuente cercana de agua? *1. Si:* *2.No:*

Si la respuesta anterior fue Si, responda:
3.9.1. ¿Qué sistema de riego utiliza? *1. Inundación* *2. Aspersión o Pivote Central:* *3. Goteo:* *4. Microasperción:*

Costos

3.10. ¿Cuánto gastó en la producción de este ciclo? *usd/ha*

Este gasto incluye: Semilla Fertilizantes Agroquímicos Arriendo Terreno Mecanización Mano de Obra

Rendimientos

3.11. ¿Cuál es el Rendimiento Esperado? *tm/ha*

3.12. ¿Cuál considera es el principal problema que afectó su rendimiento? *1. Falta de Agua:* *2. Inundaciones:* *4. Plagas /enfermedades:* *3. Fuertes vientos:* *5. Malezas:*

Si la respuesta anterior fue plagas/ enfermedad,
3.12.1 Escoga uno: *1. Roya (Puccinia polysora):* *2.Tizón Foliar (Helminthosporium turcicum):* *3. Complejo Mancha de Asfalto (Monographelo maydis y Phyllachora maydis):* *3. Gusano cogollero (Spodaptera frugiperda):* *4.Barrenador del tallo (Diatraea spp.):* *5. Mancha foliar (Curvularia lunata):*

Sección 4. Variables tomadas en Campo

4.1. ¿Qué distancia existe entre surco? *cm*

4.2. ¿Cuántas plantas existen en 10 m lineales?

4.3. ¿Cuántas mazorcas existen en 10 m lineales?

4.4. ¿Cuántos granos existen por cada mazorca?

Mazorca 1	*Mazorca 2*	*Mazorca 3*	*Mazorca 4*	*Mazorca 5*	*Mazorca 6*

4.5. ¿Cuánto pesa la mazorca (granos de la mazorca)?

Mazorca 1	*Mazorca 2*	*Mazorca 3*	*Mazorca 4*	*Mazorca 5*	*Mazorca 6*
gr.	*gr.*	*gr.*	*gr.*	*gr.*	*gr.*

4.6. ¿Cuál es el porcentaje (%) de humedad e impurezas de la muestra trillada? % *Humedad* % *Impurezas*

Sección 5. Observaciones

Nota: *El número de observación se codificará utilizando la primera letra de la siguiente manera: provincia/cantón/investigador/mes/día/No. Observación* ***Ejemplo: si la muestra fue obtenida en: Manabí/Tosagua/Gustavo Mera/Junio/05/#3; la codificación seria: MTG0605-3***

Bibliografía

[1] Anderson,T. W., Darling, D. A. *A Test of Goodness of Fit.* Journal of the American Statistical Association, Vol. 49, No. 268. (Dec., 1954), pp. 765-769.

[2] Bilgili, Mehmet. *Prediction of soil temperature using regression and artificial neural network models.* Springer-Verlag 2010.

[3] Castro, Marcelo A. *Rendimiento de maíz duro seco en verano 2016.* Coordinación de Análisis y Procesamiento de la información. MAGAP. Quito, Ecuador.

[4] Chatterjee, S., Hadi, A. *Regression Analysis by Example.* John Wiley & Sons, Inc. Fourth edition. 2006.

[5] Ciaburro, G., Venkateswaran, B. *Neural Networks with R.* Packt Publishing. 2017.

[6] Coordinación General del Sistema de información Nacional. *Manual del Investigador. Levantamiento de información para el cálculo de Rendimientos Objetivos.* Ciclo Invierno 2018.

[7] Coordinación General del Sistema de información Nacional. *Maíz Duro Seco. Boletín Situacional 2013.* MAGAP.

[8] D'Agostino, R.B., Stephens, M. A. *Goodness of Fit Tecniques.* Statistics, textbooks and monographs, vol. 68. Marcel Dekker, Inc. 1986.

[9] Echeverri, L., Arroyave,E., Barajas, F. *Comparación De Pruebas De Normalidad.* XXV Simposio Internacional de Estadística 2015.

[10] Frees, E. *Regression Modeling with Actuarial and Financial Applications.* Cambridge University Press, New York. 2010.

[11] Fisher R. A. *The use of multiple measurements in taxonomic problems.* Annual Eugenics 7(Part II): 179188.

[12] Gujarati D., Porter D. *Econometría.* McGraw-Hill. Quinta Edición. 2010.

[13] Guojun Gan, Chaoqun Ma, Jianhong Wu. *Data Clustering Theory, Algorithms, and Applications.* American Statistical Association and the Society for Industrial and Applied Mathematics. 2007.

[14] Greene, William H. *Econometric Analysis.* Prentice Hall. 7th ed.2012.

[15] Hagan, M. *et al. Neural Network Design.* Second Edition.

[16] Hebb, D. O. *The organization of behavior: A neuropsychological theory.* New York, Wiley, 1949.

[17] Hertz, J., Krogh, A., Palmer, R. *Introduction to the Theory of Neural Computation.* Lecture Notes Volume I. Santa. Fe Institute Studies in The Sciences of Complexity. Taylor & Francis Group. 1991.

[18] Hyvärinen A., Oja E. *Independent Component Analysis: A Tutorial.* http://www.cis.hut.fi/? aapo/ps/NN00.pdf. Espoo, Finland, Helsinki University of Technology.

[19] Hotelling, H. *Analysis of a complex of statistical variables into principal components.* Journal of Educational Psychology 24: 417?441 and 498?520.

[20] INEC. *Encuesta de Superficie y Producción Agropecuaria Continua (ESPAC) 2002-2016.*

[21] INIAP. *XII Reunión de maiceros de la zona andina. Memorias.* Quito. 1986.

[22] Jarque, Carlos M., Bera, Anil K. *A Test for Normality of Observations and Regression Residuals.* Source: International Statistical Review / Revue Internationale de Statistique, Vol. 55, No. 2 (Aug., 1987), pp. 163-172.

[23] Kasiviswanathan, K.S., Sudheer, K.P. and He J. *Quantification of Prediction Uncertainty in Artificial Neural Network Models.* Artificial Neural Network Modelling, Studies in Computational Intelligence 628. Springer International Publishing Switzerland 2016.

[24] Matsumura, K., *el at. Maize yield forecasting by linear regression and artificial neural networks in Jilin, China.* Journal of Agricultural Science, Page 1 of 12. Cambridge University Press 2014.

[25] Medina, P., Velín , M. *Cálculo y Análisis de Indicadores Sectoriales de Comercio Exterior para el Caso Ecuatoriano.* INEC, 2011.

[26] Minsky, M., Papert, S. *Perceptrons: An Introduction to Computational Geometry.* Cambridge, Massachusetts, USA: The MIT Press. 1969.

[27] Monteros Guerrero, A., Sarauz, Salvador. *Rendimientos de maíz duro seco en el Ecuador verano 2014.* Coordinación de Análisis y Procesamiento de la información. MAGAP. Quito, Ecuador.

[28] Monteros Guerrero, A.,Sumba Lusero, E.,Salvador Sarauz, S. *Productividad Agrícola en el Ecuador.* Coordinación de Análisis y Procesamiento de la información. MAGAP. Quito, Ecuador.

[29] Montgomery D., Peck E., Vining G. *Introducción al Análisis de Regresión Lineal.* John Wiley & Sons. Primera edición. 2002.

[30] Moreno Aguirre, B., Sarauz, S. *Determinantes del rendimiento del cultivo de maíz duro seco.* Coordinación General del Sistema de Información Nacional. MAGAP. Quito, Ecuador. 2015.

[31] National Corn Growers Association, 2018. www.worldofcorn.com

[32] Niazian, M. *et al. Modeling the seed yield of Ajowan (Trachyspermum ammi L.) using artificial neural network and multiple linear regression models.* Industrial Crops & Products 117, 224-234. Elsevier B.V. 2018.

[33] Novales A.*Econometría.* McGraw-Hill/Interamericana de España. Segunda edición. 1993.

[34] Nunes Da Silva, I., Andrade, R., Bartocci, L., Dos Reis, S. *Artificial Neural Networks, A Practical Course.* Springer, 2017.

[35] Peña, D. *Análisis de Datos Multivariantes.* 2002.

[36] Pequeno Reis, L. *et al. Prognosis on the diameter of individual trees on the eastern region of the amazon using artificial neural networks.* Forest Ecology and Management 382 (2016) 161?167. ELSEVIER. 2016.

[37] Pequeno Reis, L. *et al. Estimation of mortality and survival of individual trees after harvesting wood using artificial neural networks in the amazon rain forest.* Forest Ecology and Management 382 (2016) 161?167. ELSEVIER. 2018.

[38] Priddy, Kevin L., Keller, Paul E. *Artificial Neural Networks An Introduction.* SPIE - The International Society for Optical Engineering, Volume TT68. USA 2005.

[39] Russell, D., James G. M. *Econometric Theory and Methods.* Oxford University Press, USA. 2003.

[40] Schroeter, Benjamin J.E. *Artificial Neural Networks in Precipitation Nowcasting: An Australian Case Study.* Artificial Neural Network Modelling, Studies in Computational Intelligence 628. Springer International Publishing Switzerland. 2016.

[41] Shanmuganathan, S., Samarasinghe, S. *Artificial Neural Network Modelling.* Springer International Publishing Switzerland. 2016

[42] Segerer, C., Villodas, R. *Hidrología I.* Universidad Nacional de Cuyo. 2006.

[43] Shapiro, S., Wilk, M. B. *An Analysis of Variance Test for Normality (Complete Samples).* Biometrika, Vol. 52, No. 3/4. (Dec., 1965), pp. 591-611.

[44] Simar L., Hardler W. *Applied Multivariate Statistical Analysis.* Springer. Tercera edición. 2012.

[45] Sumba, Lucero. *Producción histórica de maíz duro seco.* Coordinación de Análisis y Procesamiento de la información. MAGAP. Quito, Ecuador.

[46] Suykens, J. *et al. Artificial Neural Networks for Modelling and Control of Non-Linear Systems.* Springer Science+Business Media Dordrecht. 1996.

[47] Üreyen, M., Gürkan, P. *Comparison of Artificial Neural Network and Linear Regression Models for Prediction of Ring Spun Yarn Properties. I. Prediction of Yarn Tensile Properties.* Fibers and Polymers 2008, Vol.9, No.1, 87-91.

[48] Wooldridge J.M. *Econometric Analysis of Cross Section and Panel Data.* The MIT Press, England.

[49] Wooldridge J.M. *Introducción a la Econometría, Un enfoque moderno.* Cengage Learning Editores. Cuarta Edición. 2010.