Instituto de Estadística
Ingeniería en Estadística y Ciencia de Datos

Clasificación de la fuga de clientes de una empresa bancaria mediante redes neuronales

Ricardo Cristian Menares Eyzaguirre

Valparaíso, Chile
8 de Julio de 2022

Índice general

Capítulo 1

Introducción

Uno de los principales problemas que se enfrentan las empresas es al de fuga de clientes, debido a que influye en la pérdida directa de los ingresos de una institución. Además, al momento de reemplazar un cliente que se desvinculó de una empresa, hay que invertir en *marketing* para poder atraer nuevos usuarios, por lo que es notorio que es más costoso y difícil obtener nuevos clientes que, por el contrario, retener a los clientes que podrían irse.

Sumado a lo anterior, los clientes resultan ser muy importantes para toda empresa, pues son el principal sustento financiero que posee. En este contexto, uno de los objetivos que debe tener una compañía es el de poder conocer a sus clientes y saber evaluar si están a gusto con los servicios de la empresa, o más específicamente, si estos usuarios pueden abandonar la empresa en el futuro. Lo anterior es sumamente importante, debido a que, si un usuario abandona la compañía, pierde los recursos financieros que obtenía de ese cliente, más aún, Miranda & et al. (2005) indican que los clientes son uno de los activos más importantes para una institución financiera, ya que está estrechamente relacionada con las utilidades del negocio.

"La fuga de clientes busca identificar los clientes con mayor probabilidad de renunciar a un producto, a un conjunto de producto o a la totalidad de los productos ofrecidos por una institución. La acción de evitar que un cliente se fugue es conocida como retención de clientes" (Jélvez, A.& et al. 2014).

En efecto, lo que se busca en este estudio es realizar un análisis de fuga de clientes, para luego poder retener al usuario ofreciéndole ofertas o productos. Es por ello, que los resultados de esta investigación tiene una importancia fundamental para el desarrollo económico y financiero de la empresa.

En este contexto, una empresa del sector bancario le motiva conocer un método eficiente para predecir si un usuario potencialmente pueda abandonar la empresa, a modo de poder retenerlo antes que efectúe el retiro, ya que re-

sulta ser más costoso conseguir nuevos clientes que retener a los que ya están en la empresa. Así, el objetivo principal de esta investigación es realizar una comparación y elección de los mejores clasificadores mediante redes neuronales artificiales (RNA) con distintas capas y neuronas para identificar y predecir los clientes que abandonan la empresa bancaria, según varias variables importantes que estos usuarios poseen en la empresa, tales como el puntaje crediticio que tiene, años en la empresa, saldo promedio, etc.

1.1. Objetivo general

Ajustar y seleccionar el mejor modelo RNA para la predicción de si un cliente ha abandonado la empresa bancaria o no.

1.2. Objetivos específicos

1. Examinar el modelo teórico elegido para los datos obtenidos.
2. Comparar experimentalmente diversos modelos RNA para la predicción propuestos.
3. Elegir el o los mejores modelos de RNA en términos de predicción obtenida.

Capítulo 2

Redes neuronales artificiales

Antes de describir lo que es una red neuronal artificial, es primordial entender una estructura neuronal real que se compone de un gran cuerpo celular y de fibras nerviosas, la cual a modo resumido se representa por la siguiente imágen:

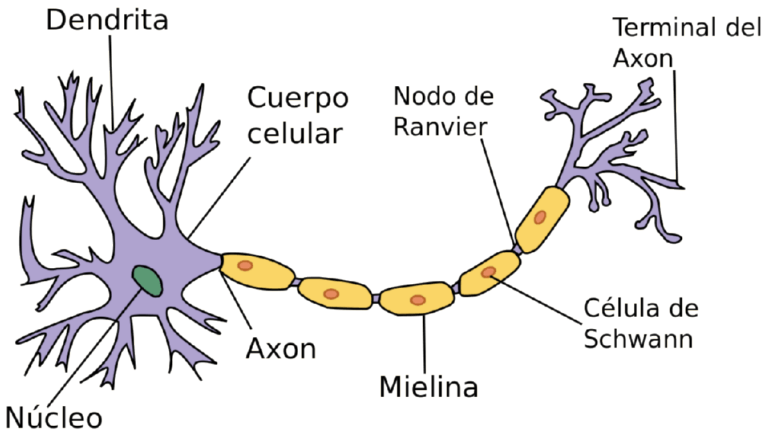

Figura 2.1: Estructura neuronal, recuperado de: `https://comofuncionaque.com/funciones-de-la-neurona/`

En esta imágen, la Dendrita cumple la misión de ser la vía de entrada (receptor) de las señales que se combinan en el cuerpo de la neurona mediante impulsos nerviosos. Luego, el axón es el camino de salida de la señal generada por la neurona. Lo que es el cuerpo amarillo de la imágen, se encuentran muchas vesículas y células que propagan señales electroquímicas de una neurona a otra. Finalmente, la neurona es estimulada por sus entradas y cuando alcanza cierto umbral, se dispara o activa pasando una señal hacia el terminal del axón en

donde se conecta con otra neurona.

En consiguiente, una red neuronal artificial (RNA) modela la relación entre un conjunto de señales de entrada y una señal de salida usando un modelo derivado desde nuestro entendimiento de cómo funciona un cerebro biológico ante estimulos externos. La forma más común de representar la estructura de una red neuronal es mediante el uso de capas (*layers*), formadas a su vez por neuronas. Cada neurona, realiza una operación sencilla y está conectada a las neuronas de la capa anterior y de la capa siguiente mediante pesos. Una imágen gráfica de una RNA es la siguiente:

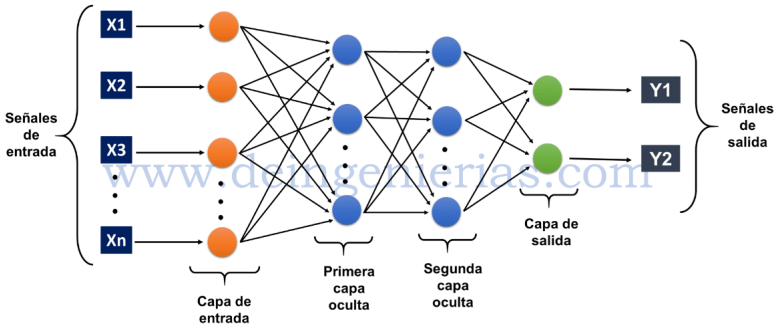

Figura 2.2: Estructura red neuronal artificial, recuperado de: `https://deingenierias.com/inteligencia-artificial/redes-neuronales-en-inteligencia-artificial/`

En la imágen, se observa que los X_i son las señales de entrada que toma esta neurona (los vectores de datos), que se almacenan directamente a una capa de entrada, para luego ir a una primera capa (notar que se mezclan los vectores) en donde se realizan funciones lineales y no lineales mediante pesos a cada observación, pasando de capa en capa conectadas hasta llegar a una capa de salida que entregará el resultado de la interacción de todas las capas del modelo en particular. Notar que en esta imágen es un modelo con 2 capas con 3 neuronas cada una, además de tener una salida binaria (puede ser una salida continua).

En consiguiente, se especifica de mejor manera las funciones que se realizan en cada neurona mediante la siguiente imágen:

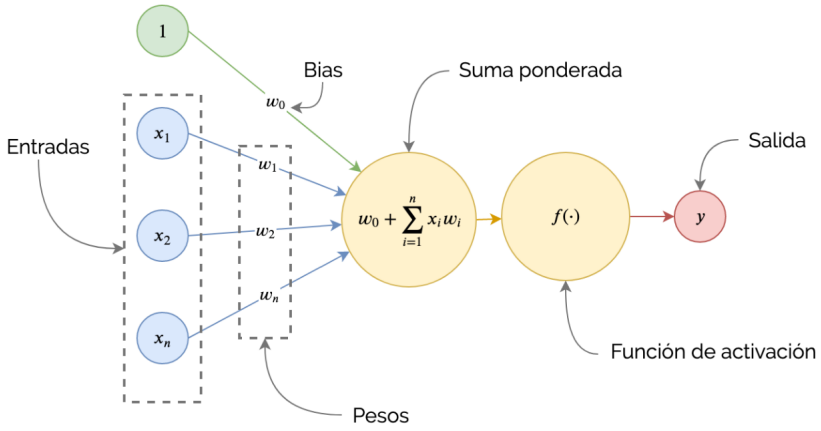

Figura 2.3: Componentes de una neurona artificial (Castillo, 2020).

Se observa que se ocupa un ̈biasq̈ue resulta ser simplemente el sesgo en específico. Para cada entrada se le asigna un peso, para luego realizar una suma ponderada de cada peso por la observación en cuestión. En consiguiente, se realiza una función de activación para esta suma ponderada obteniendo una salida. En este caso, las funciones de activación controlan en gran medida qué información se propaga desde una capa a la siguiente. Estas funciones convierten el valor neto de entrada a la neurona (combinación de los input, pesos y sesgo) en un nuevo valor. Gracias a lo anterior, estos modelos pueden aprender relaciones no lineales. Algunos de las funciones de activación más conocidas (existen muchas) son las siguientes:

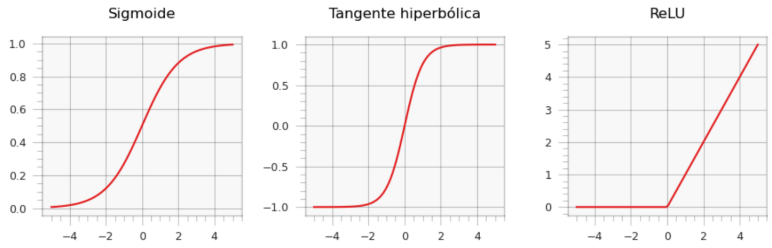

Figura 2.4: Funciones de activación más conocidas (Castillo, 2020).

Las funciones de activación de cada una, respectivamente son:

$$Sigmoide(x) = \frac{1}{1 + exp(-x)} \tag{2.1}$$

$$TanH(x) = \frac{1 - exp(-2x)}{1 + exp(-2x)} \tag{2.2}$$

$$ReLU(x) = máx(x, 0) \tag{2.3}$$

Además, se usa bastante Softplus que se considera una función continua y derivable de la función RELU, que tiene la siguiente función de activación:

$$Softplus(x) = log(1 + exp(x)) \tag{2.4}$$

Y gráficamente en comparación con RELU:

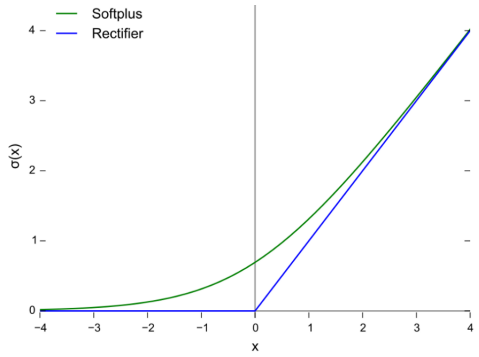

Figura 2.5: Función de activación Softplus.

Es notorio que es mucho más suave y derivable, por lo cual será la función de activación que se usará en este estudio.
Por otro lado, la función de coste es la encargada de cuantificar la distancia entre el valor real y el valor predicho por la red. Cuanto más próximo a cero es el valor de coste, mejor son las predicciones. Las más ocupadas son el error cuadrático medio(ECM) y el error medio absoluto(EMA):

$$ECM = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \tag{2.5}$$

$$EMA = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i| \tag{2.6}$$

Para terminar, se detalla cada paso que se debería realizar para hacer una red neuronal artificial:

1. Iniciar la red con valores aleatorios de los pesos y sesgo.
2. Promediar los errores de cada observación de entrenamiento al hacer su predicción.
3. Identificar la responsabilidad que ha tenido cada peso y sesgo en el error de las predicciones.
4. Modificar ligeramente los pesos y sesgos de la red (de forma proporcional a su responsabilidad en el error) en la dirección correcta para que se reduzca el error.
5. Repetir los pasos 2, 3, 4 y 5 hasta que la red sea suficientemente buena.

Capítulo 3

Metodología

3.1. Base de datos

La base de datos a usar en esta investigación cuenta con:

- 10.000 observaciones.
- 14 variables.

En donde, las variables detalladamente son:

1. **RowNumber**: enumeración del cliente 1-10.000.
2. **CustomerId**: Id del cliente.
3. **SurName**: Apellido del cliente.
4. **CreditScore**: Puntaje crediticio de varios ítems.
5. **Geography**: País de residencia. (Francia-España-Alemania).
6. **Gender**: Género.Mujer-Hombre.
7. **Age**: Edad.
8. **Tenure**: Años en la empresa. 0-8.
9. **Balance**: Saldo promedio.
10. **NumOfProducts**: N° de productos asociados a la empresa.
11. **HasCrCard**: Cliente tiene tarjeta de crédito. 0-1.
12. **IsActiveMember**: Cliente tiene cuenta activa. 0-1.
13. **EstimatedSalary**:Salario estimado en dólares estadounidenses.
14. **Exited**: Cliente se desvinculó del banco.0-1.

3.2. Análisis de datos

Se comienza buscando si existen anomalías en los datos, donde se concluye que la base de datos no cuenta con valores NaN y son los datos concordantes a sus variables.

Además, debido a la nula información que aportan para el análisis de fuga, se eliminan las 3 primeras variables:

1. RowNumber
2. CustomerId
3. SurName

Mediante el siguiente código en R:

```
1  df<-df[-c(1,2,3)]
2  dim(df) #se comprueban las dimensiones resultantes
```

Listado en R 3.1: Eliminación de variables que no aportan información relevante

En consiguiente, se visualizan gráficos de las variables de interés. Comenzando por la proporción del sexo de los clientes:

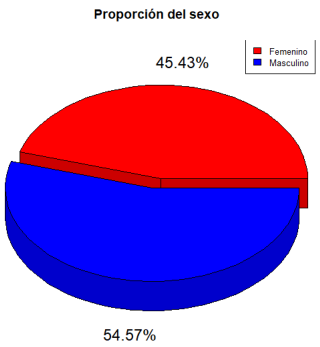

Figura 3.1: Gráfico de torta de la proporción del sexo de los clientes

Se aprecia que es casi etiquetativa esta proporción, lo cual es bueno. Ahora se presenta la proporción de clientes con tarjeta de crédito:

Figura 3.2: Gráfico de torta de la proporción de clientes con tarjeta de crédito

Es notorio que existe un desbalance de clientes con tarjetas de crédito y los que no, lo cual puede afectar a los resultados del análisis de clasificación, pues que el cliente tenga tarjeta de crédito o no puede ser un motivo importante para que decida abandonar la empresa. En consiguiente, se presenta la proporciónd e clientes activos y los que no:

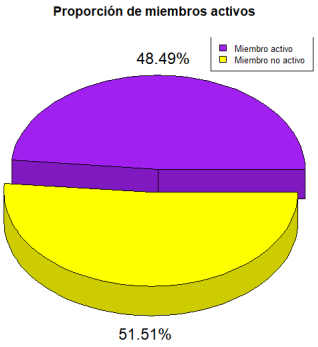

Figura 3.3: Gráfico de torta de la proporción de clientes activos

Acá si existe una proporción casi del 50 % para cada uno, lo cual es bastante bueno. Ahora se presenta la proporción de clientes que abandonan la empresa:

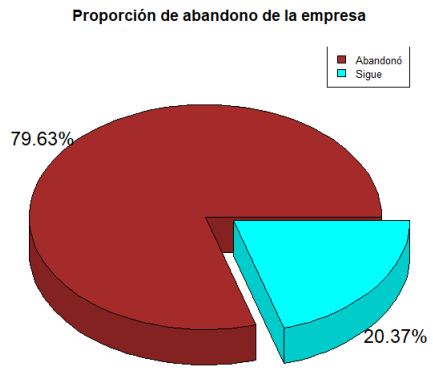

Figura 3.4: Gráfico de torta de la proporción de clientes que abandonan la empresa

Es muy evidente que existe una desproporción entre los clientes que abandonan la empresa o no, por lo cual esto puede afectar directamente a los resultados de la clasificación, pues esta variable es la que interesa clasificar. Habrá que tener cuidado con los resultados de este estudio.

Luego, se realizan gráficos de barra para las variables cualitativas con más de 2 categorías, presentando primero las frecuencias de la nacionalidad del cliente, cantidad de productos que tiene el cliente y los años de antigüedad:

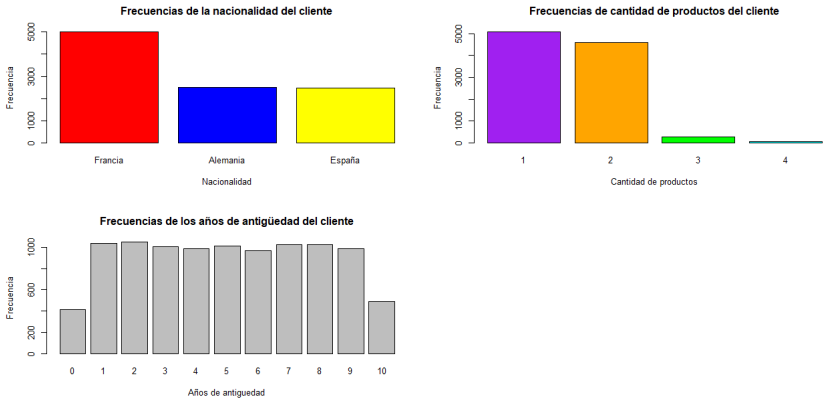

Figura 3.5: Gráficos de barras de las frecuencias de nacionalidad-cantidad de productos-antigüedad del cliente

Se aprecia que aproximadamente la mitad de los clientes tienen nacionalidad francesa y en cuanto a alemanes y españoles son equitativos. Luego, se observa que la mayoría de clientes tienen 1 o 2 productos solamente, siendo la proporción de tener más que dos productos muy baja. Por último, los años de antigüedad del cliente es bastante equitativo, a excepción de cuando tiene 0 o 10 años de antigüedad.

Ahora, se presenta un histograma y un Box-Plot del balance del cliente:

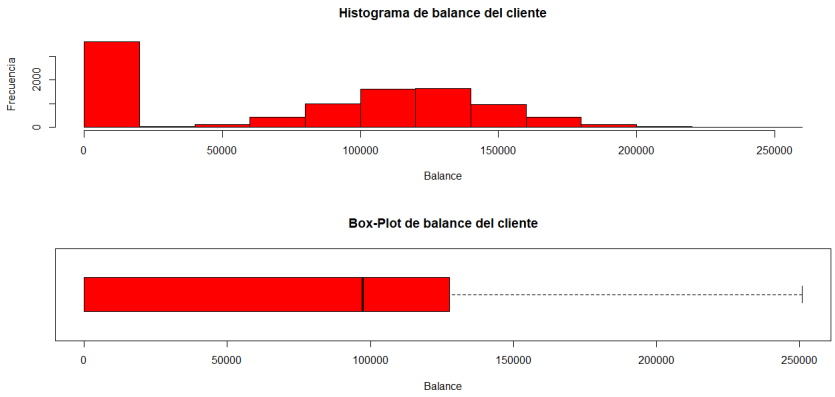

Figura 3.6: Histograma y Box-Plot del balance del cliente

Es notorio que tiene muchos valores cercanos a 0, lo cual puede intuirse un problema de colas. También se observa un comportamiento asimétrico positivo,

lo cual se explica por la cantidad de valores cercanos a 0.

Se realiza el mismo gráfico anterior pero ahora con la variable de salario estimado:

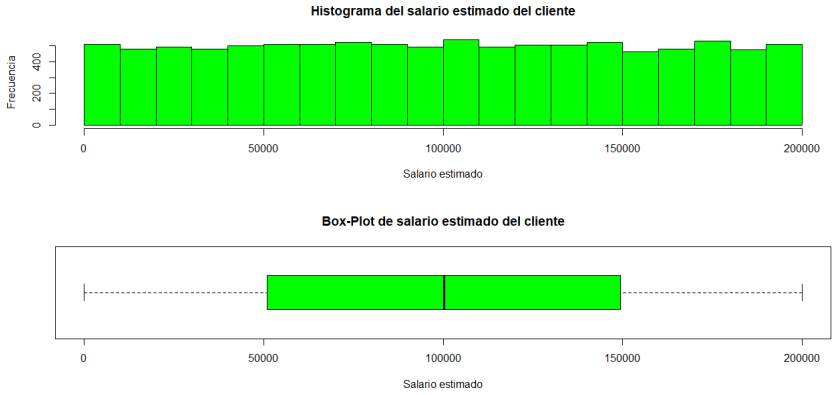

Figura 3.7: Histograma y Box-Plot del salario estimado del cliente

Se observa un comportamiento bastante equitativo, parecido al de una distribución uniforme continua. Según el Box-Plot, se observa un comportamiento simétrico

Ahora se presenta el puntaje de crédito del cliente:

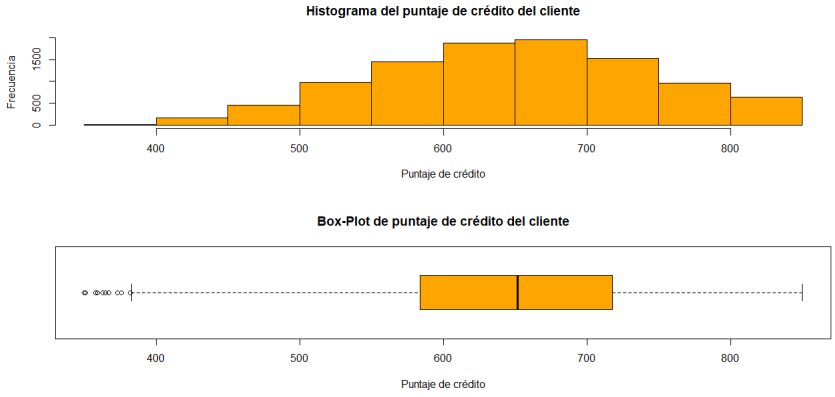

Figura 3.8: Histograma y Box-Plot del puntaje de crédito del cliente

Se observa un comportamiento simétrico y con forma de campana. Además,

se aprecian valores atípicos cuando este puntaje es menor a 350 apróx. Como no son tantos, se decide reemplazar estos valores atípicos por la media de la variable, quedando el siguiente gráfico:

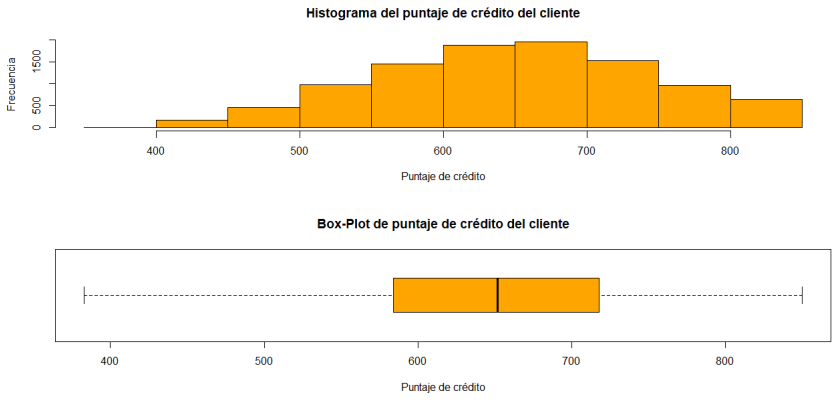

Figura 3.9: Histograma y Box-Plot del puntaje de crédito del cliente

Como se aprecia en la figura 3.9, no varío notoriamente la distribución de los datos.

Finalizando, se presenta el histograma de la edad del cliente:

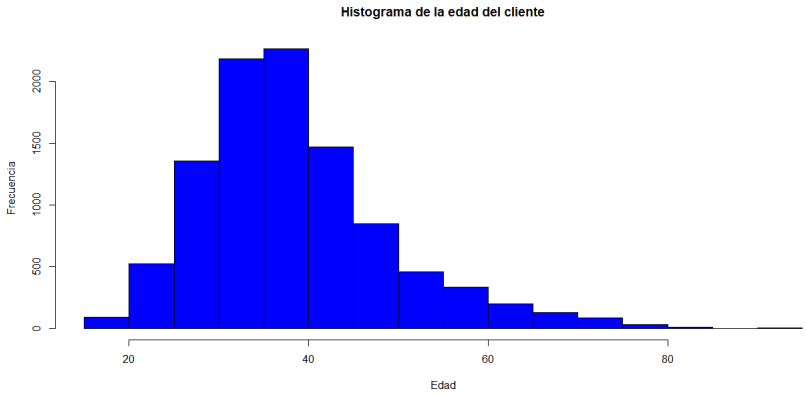

Figura 3.10: Histograma de la edad del cliente

Se aprecia que en la variable de respuesta se tiene un comportamiento de decaimiento, concentrándose mayormente en clientes entre 30 y 50 años de edad

apróx. Se observa que hay clientes bastantes viejos.

Para terminar, se presenta un gráfico de correlación para las variables cuantitativas, a modo de poder entender si existen problemas de multicolinealidad entre ellas:

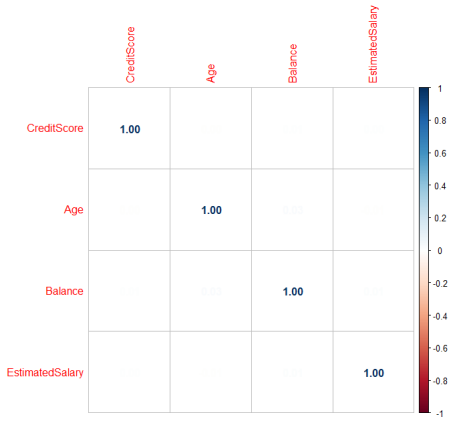

Figura 3.11: Matriz de correlaciones para las variables continuas

Se observa que las correlaciones entre las variables son muy cercanas a 0, por tanto, no ocurre el problema de que 2 o más variables estén entregando la misma información para la variable de respuesta, lo cual es bastante bueno.

3.3. Preprocesamiento de variables para RNA

A la hora de entrenar modelos basados en redes neuronales, es necesario aplicar a los datos, al menos, dos tipos de transformaciones:

1. **Binarización de variables categóricas**: crear nuevas variables dummy con cada uno de los niveles de las variables cualitativas.

2. **Estandarización y escalado de variables numéricas**: generalmente centrar los datos y estandarizar entre 0 y 1.

Entonces, para las variables categóricas se dividen en varias variables dummys, estas variables son: Geography, NumOfProducts. Se realiza mediante:

```
1 dummys<-dummy_cols(df,  select_columns = c(``Geography", ``
      NumOfProducts"))
2 names(dummys[12:18])
```

Listado en R 3.2: Creación de variables dummys

Mientras que para las variables binarias no se les realiza nada. Estas variables son: Gender, HasCrCard, IsActivedMember y Exited.

Por último, solo a las variables CreditScore, Age, Tenure, Balance y EstimatedSalary que son cuantitativas, se realiza un escalamiento estándar que ajusta los valores medidos en diferentes escalas respecto a una escala común. Básicamente, centra los datos en torno al 0 y su varianza la convierte en 1. Así, su fórmula es bastante simple, siendo:

$$X_{esc} = \frac{X_{ip} - \bar{X}_p}{S_{x_p}} \tag{3.1}$$

En R la función scale(x) realiza el escalamiento estándar a cada variable, realizando el siguiente código:

```
1 scale<-scale(df[c(1,4,5,6,10)]) #todas las variables num ricas
```

Listado en R 3.3: Escalamiento estándar de variables cuantitativas

Luego de realizar el escalamiento a los datos mediante la función de R scale(), se obtuvieron las siguientes medias y varianzas de cada variable:

Variable	**Media**	**Varianza**
CreditScore	0	1
Age	0	1
Tenure	0	1
Balance	0	1
Estimated Salary	0	1

Cuadro 3.1: Media y varianza de variables cuantitativas luego del escalamiento. Elaboración propia.

Se confirma en la tabla 3.1 las medias y varianzas que quedaron en las variables cuantitativas luego del escalamiento estándar.

3.4. Entrenamiento y prueba

Se procede a separar el conjunto de datos en dos grupos, entrenamiento y prueba. Con el primero, se entrenará los distintos modelos ANN y con el segundo se chequeará el poder predictivo que el modelo tenga. Se realiza con:

```
1  train <- createDataPartition(y = df2$Exited, p = 0.8, list = FALSE,
       times = 1)
2  df_train <- df2[train, ]
3  df_test  <- df2[-train, ]
4
5  prop.table(table(df_train$Exited))
6  prop.table(table(df_test$Exited))
```

Listado en R 3.4: Separación en entrenamiento y prueba

Capítulo 4

Resultados

4.1. Ajuste modelos RNA

Se presenta el gráfico del ajuste del modelo RNA con 1 capa y 1 neurona:

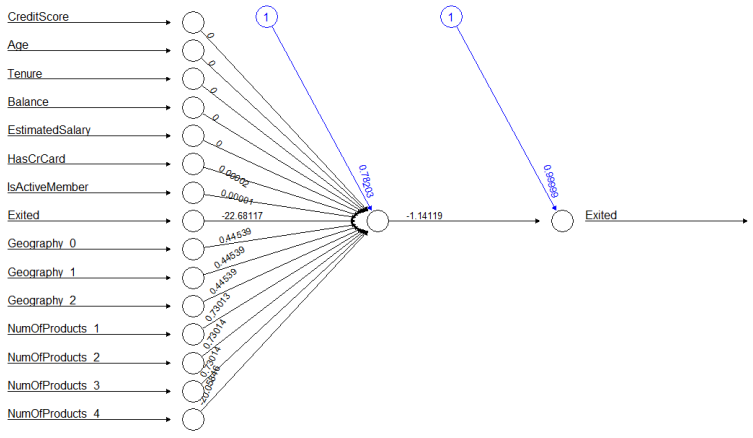

Figura 4.1: Modelo ANN con 1 capa y 1 neurona

Ahora el modelo RNA con 1 capa y 2 neuronas:

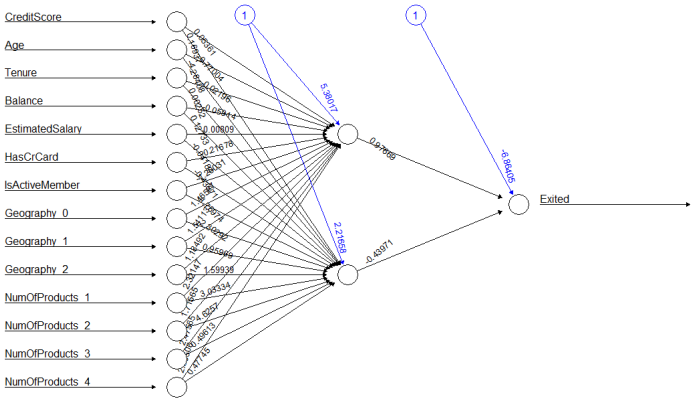

Figura 4.2: ANN de modelo con 2 capas y función de activación Softplus

En consiguiente, el modelo RNA con 2 capas y 2x3 neuronas:

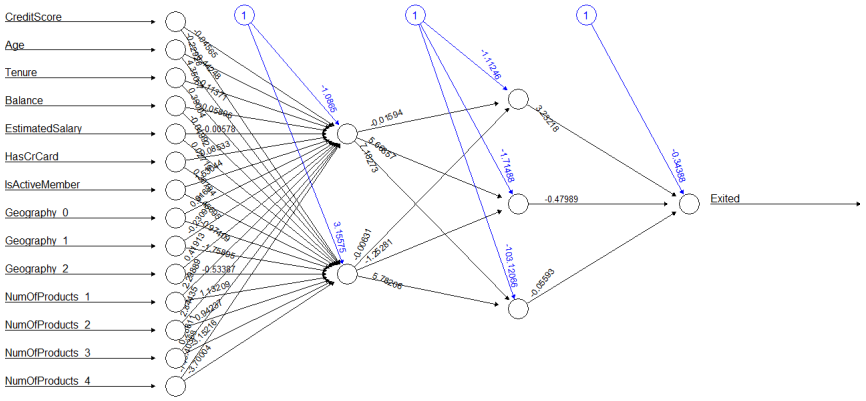

Figura 4.3: ANN de modelo con 2 capas y 2x3 neuronas con función de activación Softplus

Ahora, el modelo con 3 capas y 2x2x2 neuronas:

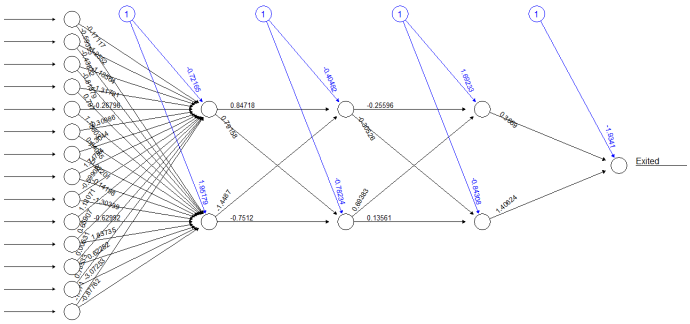

Figura 4.4: ANN de modelo con 3 capas y 2x2x2 neuronas con función de activación Softplus

Por último, el modelo con 3 capas y 4x2x3 neuronas:

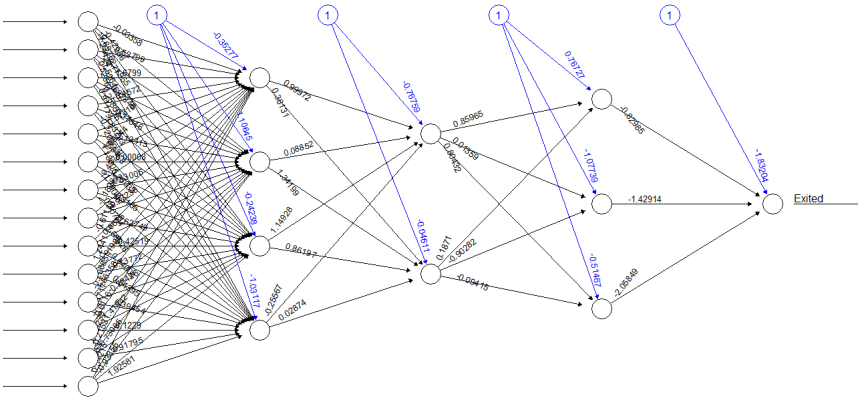

Figura 4.5: ANN de modelo con 2 capas y 2x3 neuronas con función de activación Softplus

4.2. Métricas de desempeño

Se presenta el accuracy y el error cuadrático medio de cada modelo propuesto:

Modelo RNA	**Accuracy**	**ECM**
RNA 1 capa y 1 neurona	0.849	0.151
RNA 1 capa y 2 neuronas	0.857	0.143
RNA 1 capa y 3 neuronas	0.8525	0.1475
RNA 2 capas y 2x3 neuronas	0.859	0.141
RNA 3 capas y 2x2x2 neuronas	0.874	0.126
RNA 3 capas y 4x2x3 neuronas	0.914	0.086

Cuadro 4.1: Accuracy y ECM de cada modelo RNA

Todos tuvieron buenas métricas de desempeño, en donde el modelo que más destaca es el RNA con 3 capas y 4x2x3 neuronas.

Capítulo 5

Conclusiones

Es notorio que al aumentar el número de capas, el accuracy fue aumentando. En consiguiente, el modelo mejor evaluado fue el RNA con 3 capas y 4x2x3 neuronas, por lo que el modelo escogido para poder clasificar la fuga de clientes de la empresa es este.

Cabe destacar, que existen otras técnicas de clasificación potentes que se pudieron agregar en este trabajo, tales como Random Forest, Gradient Boosting, XGBoost o SVM.

Finalizando, se destaca que a pesar de que RNA es una técnica para poder clasificar y/o predecir bastante buena, en esta se pierde capacidad de interpretación en sus resultados, debido a la gran cantidad de parámetros que se estiman, además de los distintos cálculos que se realizan. En ocasiones donde solo importa el nivel predictivo y no su interpretación, se podría recomendar probar estas técnicas para el análisis.

Capítulo 6

Referencias

- Alex. (2019). Redes neuronales en inteligencia artificial. *De ingenierías.* [Ver aquí]

- Castillo, L. (2020). Redes neuronales. *Cinvestav.* [Ver aquí]

- Graña, R. (2015). Funciones de la neurona. *cómofuncionaqué.* [Ver aquí]

- Guadarrama, E. & et al. (2015). Marketing relacional: valor, satisfacción, lealtad y retención del cliente. Análisis y reflexión teórica. *Ciencia y Sociedad, 40(2), pp. 307-340.* [Ver aquí]

- Jélvez, A. & et al. (2014). Modelo predictivo de fuga de clientes utilizando minería de datos para una empresa de telecomunicaciones en Chile. *Universidad, Ciencia y Tecnología, 18(72), pp. 100-109.* [Ver aquí]

- Latz, B. (2019). Machine Learning with R: Expert techniques for predictive modeling. 3ed. Birmingham: Packt. [Ver aquí]