

Artificial neural networks and deep learning

Hands-on project

Units 1 and 2:

Building an artificial neural network

Máster Inteligencia Artificial, UPM

Curso: 2021/2022

Fecha: 17/12/2021

Autores:

José Cerezo Egea

(jose.cerezo@alumnos.upm.es)

Jose Antonio Nicolás Navarro

(joseantonio.nicolas@alumnos.upm.es)

Arantzazu Ramos del Valle

(arantzazu.ramosd@alumnos.upm.es)

Índice de contenidos

1. Introducción.....	2
2. Proceso de diseño.....	4
3. Resultados finales	10
3.1. Hiperparámetros y modelo final	10
3.2. Pérdida y accuracy durante el entrenamiento.....	11
3.3. Resultados del conjunto de test final	11
4. Conclusiones	14
ANEXO	15

1. Introducción

El problema a resolver en esta práctica es el de clasificar jugadores de futbol en función a su calidad general como jugador basándose en sus habilidades. Para ello, se utiliza una red de neuronas artificial. El modelo neuronal se ha implementado en Python con Keras y Tensorflow en Google Collab.

El dataset inicial contiene 18207 datos y 89 atributos sobre las habilidades de futbol de los jugadores. Este dataset se limpia usando el notebook *PreparingFootballPlayersDataset.ipynb*. Tras la limpieza, el nuevo dataset contiene 17 atributos que han sido seleccionados porque tiene un valor de correlación mayor que 0.4 en valor absoluto con la puntuación global. Los atributos son: Crossing, Heading Accuracy, Short Passing, Volleys, Dribbling, Curve, FK Accuracy, Long Passing, Ball Control, Reactions, Shot Power, Stamina, Long Shots, Aggression, Positioning, Vision, and Composure. En la entrada, va a haber 16122 datos con 17 columnas (los 17 atributos) con valores entre -1 y 1, y en la salida, dará la puntuación global del jugador. En función de esta puntuación, los jugadores se clasifican en cuatro categorías; *Poor* con puntuación global entre [46, 62], *Intermediate* con puntuación global entre [63, 66], *Good* con puntuación global entre [67, 71] y *Excellent* con puntuación global entre [72, 94].

En la *figura 1*, se puede observar el formato de los atributos de entrada. Se ve como hay 17 columnas, cada una correspondiente a un atributo y que los valores de los datos de entrada varían entre -1 y 1.

Crossing	HeadingAccuracy	ShortPassing	Volleys	Dribbling	Curve	FKAccuracy	LongPassing	BallControl	Reactions	ShotPower	Stamina	LongShots	Aggression	Positioning	Vision	Composure
0.487805	-0.164557	0.260274	0.450	0.301205	0.445783	0.380952	0.297297	0.267606	0.173333	0.432099	0.188406	0.325301	0.268293	0.166667	0.121951	0.027027
0.536585	-0.367089	0.397260	0.300	0.469880	0.397590	0.166667	-0.378378	0.408451	0.306667	0.259259	-0.246377	0.036145	-0.634146	0.404762	0.414634	0.324324
-0.146341	-0.215190	0.041096	-0.350	-0.108434	-0.156627	-0.357143	0.000000	-0.154930	-0.040000	-0.234568	0.246377	-0.277108	-0.048780	0.047619	-0.024390	-0.135135
-0.317073	0.037975	0.260274	-0.025	0.132530	0.228916	-0.142857	0.270270	0.042254	0.146667	0.308642	0.826087	0.108434	0.268293	0.119048	0.121951	0.135135
0.195122	0.139241	0.123288	0.575	0.301205	-0.084337	-0.214286	-0.135135	0.239437	0.173333	0.506173	-0.101449	0.204819	-0.073171	0.428571	0.195122	0.351351
...
-0.560976	0.341772	-0.150685	-0.750	-0.325301	-0.566265	-0.642857	-0.513514	-0.408451	0.253333	-0.283951	0.246377	-0.542169	0.512195	-0.714286	-0.560976	0.081081
0.414634	-0.139241	0.506849	0.200	0.421687	0.590361	0.619048	0.324324	0.352113	0.253333	0.481481	0.101449	0.397590	-0.024390	0.285714	0.439024	0.216216
0.341463	0.063291	0.041096	-0.450	0.060241	0.084337	-0.190476	0.027027	-0.014085	0.093333	0.012346	0.536232	-0.253012	0.512195	0.119048	-0.439024	-0.162162
0.609756	0.443038	0.561644	0.925	0.542169	0.493976	0.642857	0.513514	0.521127	0.573333	0.679012	0.826087	0.710843	0.609756	0.690476	0.560976	0.540541
-0.390244	0.443038	-0.068493	-0.025	0.060241	-0.180723	-0.238095	-0.567568	0.014085	-0.066667	0.160494	-0.188406	0.132530	-0.390244	0.190476	-0.024390	-0.189189

Figura 1. Atributos de entrada.

En la *figura 2*, se puede observar los datos de salida. Estos datos están en codificación *One-hot*, todas columnas tienen valor 0 a excepción de una en el que el valor es 1 y denota a que categoría pertenece el jugador.

Poor:[46.0, 62.0]	Interm:[63.0, 66.0]	Good:[67.0, 71.0]	Excel:[72.0, 94.0]
0.0	0.0	1.0	0.0
0.0	0.0	0.0	1.0
1.0	0.0	0.0	0.0
0.0	1.0	0.0	0.0
0.0	0.0	1.0	0.0
...
0.0	0.0	1.0	0.0
0.0	0.0	1.0	0.0
0.0	1.0	0.0	0.0
0.0	0.0	0.0	1.0
0.0	1.0	0.0	0.0

Figura 2. Datos de salida.

De los datos del dataset limpio, el 80% se utilizarán para el entrenamiento de la red, el 10% para calibrar los hiperparámetros y el 10% restante para test finales para determinar la fiabilidad de la red, por tanto, quedan:

- 12897 datos para **entrenamiento**
- 1612 datos para el **test de desarrollo**
- 1613 datos para **test final**

2. Proceso de diseño

Una vez sabemos los datos con los que vamos a tratar en este problema y cuál es el objetivo de este, vamos a pasar a comentar el proceso de diseño que hemos seguido para la construcción de la red neuronal que nos resuelva dicho problema.

En primer lugar, he de comentar que vamos a seguir el proceso de diseño comentado en clase, ilustrado en la *figura 3*. Este diagrama consiste en primer lugar alcanzar un valor de bias relativamente bajo para después pasar a reducir la varianza, lo que probablemente cause que el bias vuelva a incrementar y se tenga que reducir. Este proceso se sigue de forma iterativa hasta que alcancemos una red neuronal que satisfaga nuestras condiciones.

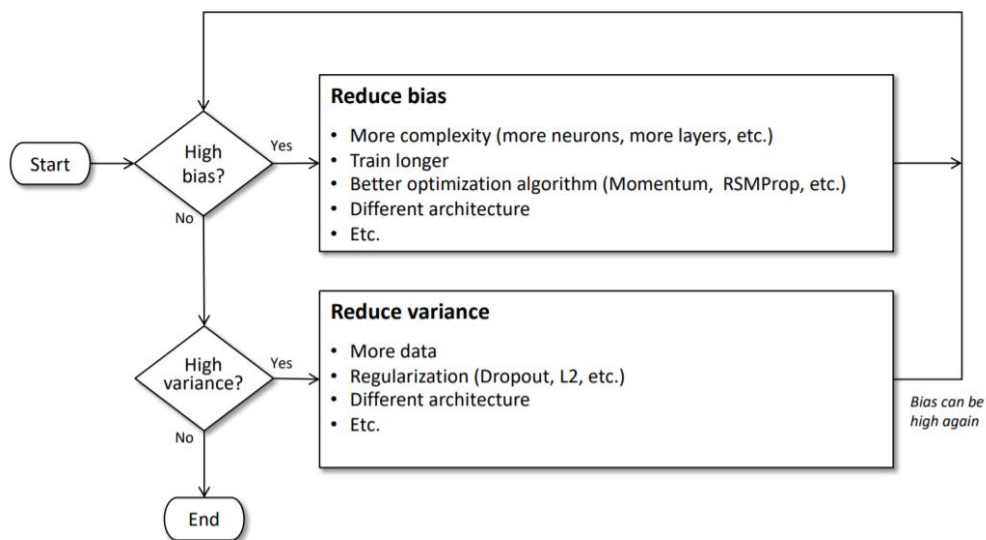


Figura 3. Proceso de diseño de una red neuronal. Tema 2.5. "Regularization", UPM. (2021). M. Molina.

En el caso de las condiciones, lo que se desea es una arquitectura neuronal que sea capaz de realizar un trabajo similar al de un humano. Para el caso de nuestro problema, hemos asumido un error humano del 10%. Este valor está basado en la creencia de que, dado un humano y presentándole las 17 características de un jugador en concreto, nos parece razonable que 1 de cada 10 veces se falle en la clasificación de este jugador.

Con el objetivo de que el proceso de diseño sea replicable, establecemos una semilla de la siguiente forma: `random.seed(10)`.

Pasando ya al proceso al proceso de diseño como tal, la primera arquitectura con la que decidimos comenzar fue una un tanto sencilla, como lo es una arquitectura neuronal con dos capas ocultas, compuestas de 25 neuronas cada una. Además, estas capas tendrán como función de activación una *sigmoidal*. En cuanto al resto de hiperparámetros, estableceremos el número de épocas a 10, el valor de *learning_rate* a 0.1, pues no podemos empezar con algo muy bajo ya que tenemos pocas épocas, y el *batch_size* a 128, pues este debe de ser una potencia de dos y además consideramos que es un valor adecuado teniendo en cuenta el número de casos de entrenamiento de los que disponemos.

Ya teniendo clara la forma que tiene la red neuronal inicial y el valor de sus parámetros, veamos los resultados que obtiene:

```
loss=0.838, categorical_accuracy=0.625, val_loss=0.827, val_categorical_accuracy=0.638
```

Como era de esperar, tanto el valor de *accuracy* en entrenamiento (0.625) como en validación (0.638) es bajo, pues la red es muy simple.

Viendo esto, pasamos a realizar nuestra primera modificación, que tiene como objetivo reducir el bias. Esta modificación consiste en aumentar la complejidad de la arquitectura neuronal, pues ahora constará de tres capas ocultas con 75, 50 y 25 neuronas, respectivamente. Veamos los nuevos resultados:

```
loss=1.26, categorical_accuracy=0.472, val_loss=1.22, val_categorical_accuracy=0.537
```

Seguimos teniendo un bias elevado, de hecho, aún más elevado que antes, puesto que hemos aumentado la complejidad de la red neuronal, pero las funciones de activación siguen siendo “básicas” y el número de épocas parece no ser suficiente. Por tanto, lo siguiente que hicimos fue cambiar la función de activación *sigmoidal* por una *ReLU* (en todas las capas ocultas). Estos son los resultados:

```
loss=0.46, categorical_accuracy=0.796, val_loss=0.427, val_categorical_accuracy=0.821
```

En este caso sí que hemos conseguido reducir drásticamente el bias, pues como hemos visto en teoría, la función de activación ReLU nos aporta una serie de ventajas frente a otras como la sigmoideal o tangente hiperbólica. Ahora, haremos la otra modificación que consideramos más necesaria, aumentar el número de épocas a 100, pues probablemente sean insuficientes. Los resultados obtenidos son los siguientes:

```
loss=0.37, categorical_accuracy=0.839, val_loss=0.43, val_categorical_accuracy=0.823
```

Como era de esperar el bias ha sido reducido pues le hemos dado más tiempo a la red neuronal para aprender. Aun así, la red parece necesitar algo más de tiempo para aprender, pues los valores de *accuracy* no oscilan, sino que no hacen más que incrementar, lo cual es una señal de que necesitamos un mayor número de épocas. En este caso empleamos 500 épocas y estos son los resultados:

```
loss=0.169, categorical_accuracy=0.933, val_loss=0.967, val_categorical_accuracy=0.765
```

De nuevo, hemos reducido el bias de la red, que ahora tiene un valor negativo, pues el error humano que estamos teniendo en cuenta es incluso mayor que el cometido por la red en entrenamiento. Además, hemos observado que durante el paso de las épocas los valores de *accuracy* ya no crecían de forma lineal y constante como en el caso anterior, sino que se percibían ciertas oscilaciones, lo que indica que subir más las épocas sin realizar cambios previos no tendría mucho sentido.

A lo que hemos de prestar intención es al valor de bias, pues al ser negativo y teniendo en cuenta el proceso de diseño de la figura 3, debemos de pasar ahora a reducir el valor de la varianza del modelo.

Para conseguir esto, lo que hicimos fue introducir a nuestra red un regularizador ya que este permitirá que los pesos de la red (w) no crezcan demasiado y por tanto no se ajusten en exceso (sobreentrenen) a los datos de entrada. En concreto, el regularizador que empleamos es L2 con un factor de regularización (λ) de 0.001 (en todas las capas ocultas y en la capa de salida). Decidimos emplear este y no L1 por su capacidad de lograr mantener los pesos más pequeños, ya que penaliza con estos al cuadrado. Los resultados que obtuvimos fueron los siguientes:

```
loss=0.477, categorical_accuracy=0.816, val_loss=0.473, val_categorical_accuracy=0.832
```

En efecto, el introducir un regularizador hizo que la red no se ajustara tanto a los datos de entrenamiento y por tanto consiguiera generalizar de una mejor forma, lo que lleva a unos mejores resultados sobre los datos de desarrollo (menor varianza), con el inconveniente de que el bias aumenta.

Una vez llegamos a estos resultados, estuvimos realizando una serie de pruebas para ver si conseguíamos reducir el bias sin perder mucha varianza o, al contrario, reducir varianza sin que afectara demasiado al bias. Estas pruebas se compusieron principalmente de cambiar la arquitectura neuronal, tanto número de capas ocultas como de neuronas en esas capas, así como cambiar el número de épocas.

Haciendo esto, observamos que la medida de *accuracy* producía oscilaciones, es decir, a partir de una arquitectura neuronal lo suficientemente compleja como para tratar el problema, las modificaciones que añadíamos no mejoraban los resultados, pues siempre se oscilaba en torno al 82% de precisión en los datos de desarrollo. Por tanto, lo que decidimos fue mantener la arquitectura neuronal que se ha comentado anteriormente y realizamos una serie de pruebas sobre la misma, variando diferentes parámetros con el objetivo de tratar de superar este último valor de *accuracy*, 83.2%.

La primera de las pruebas que hicimos se corresponde con el uso de optimizadores alternativos al que habíamos venido utilizando hasta este momento, con el objetivo de mejorar el bias consiguiendo una precisión mayor con los datos de entrenamiento. Este uso alternativo se fundamenta en el hecho comentado en la asignatura que destaca que cada uno de estos optimizadores puede resultar ser mejor que otro dependiendo del problema que se quiere afrontar. Los optimizadores alternativos que consideramos fueron *Adam*, *SGD* con la variación de *Nesterov (NAG)* y *RMSprop*.

Para el primero de los optimizadores, *Adam*, establecimos los hiperparámetros con los valores predefinidos que concreta la librería *Keras* (0.9 para el *momentum decay* y 0.999 para el *scaling decay*) sin alterar la tasa de aprendizaje y se obtuvieron los resultados que se muestran a continuación:

```
loss=0.559, categorical_accuracy=0.776, val_loss=0.545, val_categorical_accuracy=0.806
```

Se puede observar que el uso de este optimizador empeora la precisión con respecto a los datos de entrenamiento, no consiguiendo mejorar el bias que era lo que se buscaba, por lo que descartamos su uso de forma inmediata.

El siguiente optimizador con el que probamos fue *NAG*. De nuevo, no se varió el valor de la tasa de aprendizaje y el valor del factor *momentum* se estableció en 0.9 como se aconseja en la asignatura. Los resultados obtenidos fueron en la línea del optimizador anterior

```
loss=0.508, categorical_accuracy=0.805, val_loss=0.529, val_categorical_accuracy=0.803
```

Donde podemos ver que, si bien el bias mejora ligeramente con respecto a *Adam*, no lo consigue hacer con respecto a los mejores resultados obtenidos hasta el momento. Así que, al no poder reducir el bias, descartamos el uso de este optimizador en pruebas futuras.

El tercero de los optimizadores que comprobamos para la mejora del bias es *RMSprop*. Como viene siendo habitual, el valor de la tasa de aprendizaje se mantiene como hasta ahora y el valor para el *decay rate* se establece en 0.9.

```
loss=0.727, categorical_accuracy=0.72, val_loss=0.753, val_categorical_accuracy=0.731
```

Se puede ver que, de los tres optimizadores con los que se ha experimentado, es el que peor desempeño realiza y el que más bias genera, que es lo contrario a lo que se quería conseguir. Es por ello que, en las posteriores etapas de la experimentación realizada, se volverá al optimizador original SGD.

Las pruebas que se han llevado a cabo a continuación tienen que ver con el uso de funciones de activación diferentes a *ReLU* en las capas ocultas. Esta experimentación no tiene como objetivo reducir el *bias* o la varianza en concreto, sino intentar mejorar el rendimiento general de la red neuronal construida, tratando así con el problema de *vanishing gradient* que podría estar dándose, ya que la red converge rápidamente y el valor de precisión con los datos de entrenamiento se estanca fácilmente. Esto podría deberse también por los datos que se proporcionan en la práctica, pero consideramos que es una prueba con la que no se tiene nada que perder y existe un potencial beneficio general en el desempeño de la red neuronal.

Al igual que con las pruebas anteriores, se van a considerar tres funciones de activación alternativas diferentes; *ELU*, *Leaky ReLU* y *SeLU*. Los resultados se muestran a continuación y cada listado de valores se corresponde con las funciones de activación que tenemos en cuenta respectivamente:

```
loss=0.494, categorical_accuracy=0.808, val_loss=0.479, val_categorical_accuracy=0.829
```

```
loss=0.482, categorical_accuracy=0.814, val_loss=0.516, val_categorical_accuracy=0.81
```

```
loss=0.477, categorical_accuracy=0.816, val_loss=0.472, val_categorical_accuracy=0.829
```


Lo más destacable de los resultados obtenidos es que se obtienen unas métricas muy similares a las obtenidas con la función de activación *ReLU* y apenas tienen impacto en el bias y la varianza, por lo que decidimos no cambiar la función *ReLU* por ninguna de las nombradas.

La siguiente prueba tiene que ver con la experimentación de distintos valores para la tasa de aprendizaje. Con esto lo que se pretende es comprobar qué valor resulta más efectivo para que la red neuronal aprenda de forma consistente y razonablemente rápido sin que se escape demasiado de los óptimos donde cae. Así que probamos con los valores 0.01, y 0.001. Para este último valor realizamos una prueba adicional aumentando el número de épocas a 1000 para que, aunque la red se entrenara lentamente, tuviera margen suficiente de mejora. Los resultados obtenidos son los siguientes, para cada prueba respectivamente:

```
loss=0.471, categorical_accuracy=0.83, val_loss=0.473, val_categorical_accuracy=0.837
```

```
loss=0.595, categorical_accuracy=0.799, val_loss=0.585, val_categorical_accuracy=0.819
```

```
loss=0.553, categorical_accuracy=0.812, val_loss=0.548, val_categorical_accuracy=0.822
```

Existen dos hechos destacables en los resultados obtenidos. El primero es que, para el valor de 0.01 se ha conseguido mejorar tanto el bias como la varianza, ligeramente, pero de forma suficientemente sustancial para establecerlo como nuevo valor para la tasa de aprendizaje. El segundo hecho destacable es que la prueba con más épocas y un valor más bajo para la tasa de aprendizaje no arroja mejores resultados, por lo que se va a mantener el número de épocas en 500. Además, durante el entrenamiento, el valor de precisión para los datos de entrenamiento y validación iba oscilando mucho, pero no mejoraba.

Continuamos la experimentación modificando el tamaño de *batch* con un valor superior e inferior a 128, que es el que veníamos manejando hasta ahora. Con esta idea, al igual que con la prueba anterior, lo que se pretende es mejorar el modelo de manera global, debido al estancamiento que se comentó al inicio de estas pruebas. Los tamaños de *batch* escogidos son 64 y 256 (potencia de 2, por temas de memoria caché) y se obtuvieron los siguientes resultados, respectivamente:

```
loss=0.456, categorical_accuracy=0.831, val_loss=0.473, val_categorical_accuracy=0.829
```

```
loss=0.502, categorical_accuracy=0.822, val_loss=0.503, val_categorical_accuracy=0.831
```

Se puede comprobar que no se obtiene una mejora sustancial del bias o la varianza y concluimos que lo mejor es no alterar este valor, pues con un tamaño de 128 es con el que se ha obtenido un mejor desempeño en el modelo.

Llegados a este punto, una mejora en el bias implica un empeoramiento de la varianza, por lo que no tiene sentido intentar reducirlo nuevamente. Aun a sabiendas que de muy probablemente habíamos alcanzado un buen modelo para la práctica, decidimos seguir probando introduciendo la técnica *dropout* en las capas ocultas. Los resultados obtenidos con probabilidades del 10% y del 20% respectivamente se muestran a continuación:

```
loss=0.519, categorical_accuracy=0.806, val_loss=0.492, val_categorical_accuracy=0.828
```

```
loss=0.554, categorical_accuracy=0.794, val_loss=0.506, val_categorical_accuracy=0.829
```

Como se ve, la varianza empeora y también lo hace la precisión con los datos de entrenamiento, así que descartamos su uso, no sin antes realizar una última prueba introduciendo normalización de batch junto a dropout y L2 (una combinación de 3 regularizadores distintos) con los siguientes resultados obtenidos:

```
loss=0.542, categorical_accuracy=0.784, val_loss=0.48, val_categorical_accuracy=0.821
```

Que apenas aportan algo nuevo al resto de pruebas y se descartó su uso por esa razón.

3. Resultados finales

3.1. Hiperparámetros y modelo final

En esta sección se va a comentar los hiperparámetros que se han utilizado para el modelo final, así como este modelo.

En el modelo final, se ha utilizado una red neuronal con tres capas ocultas, en las que hay 75, 50 y 25 neuronas respectivamente. Como valor de learning rate, α , se ha utilizado 0.01.

En las capas ocultas, la función de activación que se utiliza es la ReLU. Para la salida, se utiliza una función de activación Softmax.

En el entrenamiento, se han utilizado 500 épocas y un tamaño de mini-batch de 128.

A continuación, se muestra el modelo final:

```
# MODELO
n_neurons_per_hidden_layer = [75, 50, 25]
learning_rate = 0.01

model = keras.Sequential(name="my_model")

model.add(keras.layers.InputLayer(input_shape=(INPUTS,)))
for neurons in n_neurons_per_hidden_layer:
    model.add(keras.layers.Dense(neurons, activation="relu",
                                   kernel_regularizer=keras.regularizers.l2(0.001)))
model.add(keras.layers.Dense(OUTPUTS, activation="softmax",
                               kernel_regularizer=keras.regularizers.l2(0.001)))

model.compile(loss=tf.keras.losses.categorical_crossentropy,
               optimizer=tf.keras.optimizers.SGD(learning_rate=learning_rate),
               metrics=["categorical_accuracy"])

# ENTRENAMIENTO
n_epochs = 500
batch_size = 128
start_time = time.perf_counter()

history = model.fit(x_train, t_train,
                    batch_size=batch_size,
                    epochs=n_epochs,
                    verbose=1,
                    validation_data=(x_dev, t_dev))
```

3.2. Pérdida y accuracy durante el entrenamiento

La siguiente gráfica (*figura 4*), muestra la evolución de la pérdida y la precisión a lo largo del entrenamiento. En esta se puede observar como con el paso de las épocas nuestro modelo va aprendiendo poco a poco, aunque en las épocas finales este aprendizaje no sea tan notable. Además, se puede apreciar como no estamos sobreentrenando la red, pues tanto para el valor de pérdida, como para el de precisión, las diferencias entre estas métricas en sus respectivos resultados de entrenamiento y desarrollo son mínimas (esto se puede apreciar visualmente, pues la línea roja y naranja, que representan la precisión en entrenamiento y desarrollo, respectivamente, están muy juntas, al igual que sucede con las líneas verde y azul).

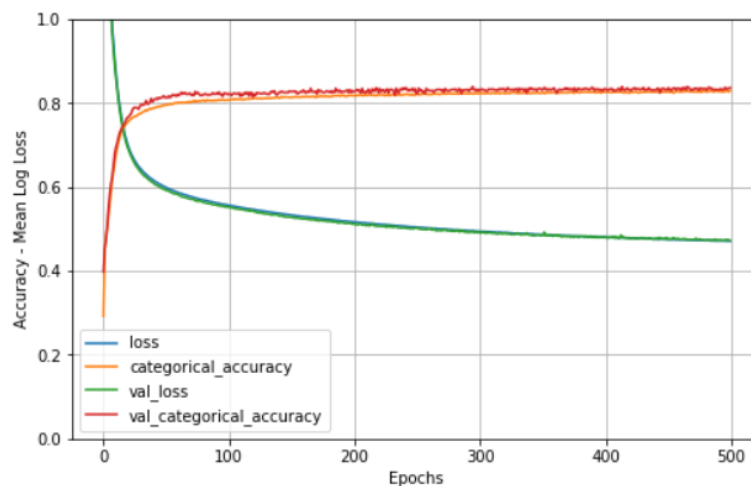


Figura 4. Evolución del accuracy y loss durante el entrenamiento.

El proceso de entrenamiento completo, en el que se muestran los valores de la pérdida (*log loss*) y de precisión (*accuracy*) así como el tiempo de computación para cada época puede encontrarse en el anexo.

3.3. Resultados del conjunto de test final

La siguiente tabla (*tabla 1*), muestra la matriz de confusión obtenida con el modelo final. Esta matriz es obtenida durante el test final, comparando los resultados predichos por la red con los resultados objetivo o reales.

		Clase predicha				
		Poor	Interm	Good	Excel	Total
Clase real	Poor	364	54	2	0	420
	Interm	56	303	56	0	415
	Good	2	55	337	35	429
	Excel	6	0	38	305	349
Total		428	412	433	340	1613

Tabla 1. Matriz de confusión.

En cada fila, se muestra cuantas veces se ha predicho cada clase para una clase real determinada. Es decir, la fila *Interm* – columna *Good*, indica el número de veces que, siendo la clase real *Interm*, se ha predicho la clase *Good*. Por ejemplo, en la fila 1, para los casos en los que la clase real es *Poor*, se ha predicho 364 veces *Poor*, 54 *Interm*, 2 *Good* y ninguna *Excel*.

Por lo tanto, en esta matriz los valores de la diagonal principal son número de aciertos y el resto son fallos. Si sumamos los números de la diagonal principal obtenemos 1309, el número total de aciertos, y 304 es el número de fallos (fallos = 1613-1309). Por tanto, obtenemos que el *accuracy* sobre los datos del subconjunto de test es de un 81.15% ($accuracy = 1309/1613$).

La tabla que se muestra a continuación (tabla 2), muestra un desglose de los valores *precisión*, *recall*, *f1-score* y *support* para cada clase.

	Precision	recall	f1-score	support
Poor	0.88	0.86	0.87	420
Interm	0.74	0.73	0.73	415
Good	0.78	0.79	0.78	429
Excel	0.90	0.87	0.89	349
micro avg	0.82	0.81	0.81	1613
macro avg	0.82	0.81	0.82	1613
weighted avg	0.82	0.81	0.82	1613
samples avg	0.81	0.81	0.81	1613

Tabla 2. Classification report.

Precision (*true positive rate*) es la proporción de observaciones correctamente predichas para una clase, frente a todas las veces que esta clase ha sido predicha. Por ejemplo, en la tabla 1 se puede ver que la clase *Good* ha sido predicha correctamente 337 veces (valor de *Good* en la diagonal principal), frente a las 433 veces que ha sido predicha en total (suma de los valores de la columna *Good*). Por lo tanto, el valor de *precision* para la clase *Good* es de 0.78 ($precision = 337/433$).

Recall (sensitivity) es la fracción de observaciones correctamente predichas de una clase. Es decir, la proporción de observaciones correctamente predichas, frente a todas las observaciones que tienen realmente esa clase. Por ejemplo, en la *tabla 1* se puede ver que la clase *Good* ha sido predicha correctamente 337 veces (valor de *Good* en la diagonal principal), frente a las 429 observaciones en las que esta es la clase real (suma de los valores de la fila *Good*). Por lo tanto, el valor de *recall* para la clase *Good* es de 0.79 ($recall = 337/429$).

F1-score es la media ponderada entre los valores de *precision* y *recall*. Puede llegar a ser más útil que el *accuracy*, sobretodo cuando la distribución entre clases no es homogénea. El *accuracy* es mejor si *precision* y *recall* tienen costes similares. El *F1-score* se ha calculado de la siguiente forma:

$$F1 - score = \frac{2 \cdot (Precision \cdot Recall)}{Precision + Recall}$$

Así que, por ejemplo, para la clase *Good*, obtenemos un *F1-score* de 0.78.

El valor *support* es cuantas observaciones reales hay de cada clase.

Para calcular los valores de *micro average* se suma todos los dividendos y divisores de cada métrica para cada clase y obtiene el cociente global. Por ejemplo, en *recall* tenemos que los dividendos-divisores de cada clase son: 364/420 *Poor*, 303/415 *Interm*, 337/429 *Good* y 305/349 *Excell*. Por tanto, el valor de *micro average* de esta métrica será 0.81 ($\frac{364+303+337+305}{420+415+429+349} = \frac{1309}{1613}$).

Para el *macro average*, se realiza una media de todas las métricas, dando igual importancia a todas las clases. Por ejemplo, en *recall* tenemos que los valores de esta métrica para cada clase son: 0.86 *Poor*, 0.73 *Interm*, 0.79 *Good* y 0.87 *Excell*. Por tanto, el valor de *macro average* de esta métrica será 0.81 ($\frac{0.86+0.73+0.79+0.87}{4}$).

En *weighted average*, se realiza la media ponderada de las métricas, en las que cada clase tiene un peso en función del número de observaciones de esa clase en los datos reales. Por ejemplo, en *recall* tenemos que los valores de esta métrica y las observaciones reales para cada clase son: 0.86 – 420 *Poor*, 0.73 – 415 *Interm*, 0.79 – 429 *Good* y 0.87 – 349 *Excell*. Por tanto, el valor de *weighted average* de esta métrica será 0.81 ($\frac{0.86 \cdot 420 + 0.73 \cdot 415 + 0.79 \cdot 429 + 0.87 \cdot 349}{4}$).

Para *samples average*, calcula la métrica directamente de los valores totales. Por ejemplo, para *precision*, hay un total de 1309 valores acertados frente a las 1613 observaciones totales, por lo que el *samples average* de *precision* será 0.81 (1309/1613).

4. Conclusiones

Hemos observado que los mejores resultados se obtienen usando tres capas ocultas con 75, 50 y 25 neuronas ocultas en cada una de ellas y con un learning rate de 0.01.

Tras varias pruebas, creemos que no es posible superar el 84% de *accuracy* en el desarrollo, pues hemos trabajado con una amplia cantidad de modelos y ninguno supera esta medida, es más, todos acaban oscilando en ese valor. Esto creemos que puede venir dado por dos motivos, el primer es por la complejidad del problema, pues no es trivial y el segundo sería el preprocesado de datos que se realizó, pues se eliminaron una gran cantidad de atributos que, aunque su correlación con la variable de salida fuera inferior a los que utilizamos en este problema, seguramente siguen siendo relevantes. Por tanto, formas de mejorar estos resultados podrían ser la de emplear una mayor cantidad de atributos (parámetros de entrada) o aumentar el número de datos.

Se puede observar como la red neuronal tiene unos mejores resultados a la hora de predecir las clases extremos (*Poor* y *Excellent*) y peor para los intermedios (*Intermediate* y *Good*), siendo esto lo común también si empleáramos esta tarea a un humano, pues es más fácil simplificar e ir a los extremos, que diferenciar entre si un jugador es intermedio o bueno, pues la línea que separa estos conceptos es más compleja. Aun así, se obtienen buenas métricas en general.

ANEXO

Este anexo contiene el proceso de aprendizaje completo, época a época, del modelo final comentado en el apartado 3.

```
Epoch 1/500
101/101 [=====] - 1s 4ms/step - loss: 1.4882 - categorical_accuracy: 0.2923 - val_loss: 1.4579 - val_categorical_accuracy: 0.3983
Epoch 2/500
101/101 [=====] - 0s 2ms/step - loss: 1.4141 - categorical_accuracy: 0.4514 - val_loss: 1.3759 - val_categorical_accuracy: 0.4516
Epoch 3/500
101/101 [=====] - 0s 2ms/step - loss: 1.3218 - categorical_accuracy: 0.4809 - val_loss: 1.2806 - val_categorical_accuracy: 0.4721
Epoch 4/500
101/101 [=====] - 0s 2ms/step - loss: 1.2298 - categorical_accuracy: 0.5027 - val_loss: 1.1962 - val_categorical_accuracy: 0.5056
Epoch 5/500
101/101 [=====] - 0s 2ms/step - loss: 1.1532 - categorical_accuracy: 0.5299 - val_loss: 1.1259 - val_categorical_accuracy: 0.5484
Epoch 6/500
101/101 [=====] - 0s 2ms/step - loss: 1.0889 - categorical_accuracy: 0.5579 - val_loss: 1.0657 - val_categorical_accuracy: 0.5837
Epoch 7/500
101/101 [=====] - 0s 3ms/step - loss: 1.0325 - categorical_accuracy: 0.5870 - val_loss: 1.0130 - val_categorical_accuracy: 0.6086
Epoch 8/500
101/101 [=====] - 0s 2ms/step - loss: 0.9828 - categorical_accuracy: 0.6156 - val_loss: 0.9665 - val_categorical_accuracy: 0.6247
Epoch 9/500
101/101 [=====] - 0s 2ms/step - loss: 0.9395 - categorical_accuracy: 0.6387 - val_loss: 0.9261 - val_categorical_accuracy: 0.6532
Epoch 10/500
101/101 [=====] - 0s 2ms/step - loss: 0.9018 - categorical_accuracy: 0.6586 - val_loss: 0.8906 - val_categorical_accuracy: 0.6799
Epoch 11/500
101/101 [=====] - 0s 2ms/step - loss: 0.8688 - categorical_accuracy: 0.6785 - val_loss: 0.8582 - val_categorical_accuracy: 0.6960
Epoch 12/500
101/101 [=====] - 0s 2ms/step - loss: 0.8394 - categorical_accuracy: 0.6939 - val_loss: 0.8302 - val_categorical_accuracy: 0.7060
Epoch 13/500
101/101 [=====] - 0s 2ms/step - loss: 0.8135 - categorical_accuracy: 0.7075 - val_loss: 0.8046 - val_categorical_accuracy: 0.7196
Epoch 14/500
101/101 [=====] - 0s 2ms/step - loss: 0.7909 - categorical_accuracy: 0.7192 - val_loss: 0.7825 - val_categorical_accuracy: 0.7301
Epoch 15/500
101/101 [=====] - 0s 2ms/step - loss: 0.7710 - categorical_accuracy: 0.7299 - val_loss: 0.7632 - val_categorical_accuracy: 0.7364
Epoch 16/500
101/101 [=====] - 0s 2ms/step - loss: 0.7537 - categorical_accuracy: 0.7333 - val_loss: 0.7459 - val_categorical_accuracy: 0.7444
Epoch 17/500
101/101 [=====] - 0s 2ms/step - loss: 0.7386 - categorical_accuracy: 0.7393 - val_loss: 0.7311 - val_categorical_accuracy: 0.7488
Epoch 18/500
101/101 [=====] - 0s 2ms/step - loss: 0.7253 - categorical_accuracy: 0.7458 - val_loss: 0.7180 - val_categorical_accuracy: 0.7525
Epoch 19/500
101/101 [=====] - 0s 2ms/step - loss: 0.7138 - categorical_accuracy: 0.7509 - val_loss: 0.7064 - val_categorical_accuracy: 0.7605
Epoch 20/500
101/101 [=====] - 0s 2ms/step - loss: 0.7038 - categorical_accuracy: 0.7535 - val_loss: 0.6966 - val_categorical_accuracy: 0.7655
Epoch 21/500
101/101 [=====] - 0s 2ms/step - loss: 0.6947 - categorical_accuracy: 0.7575 - val_loss: 0.6883 - val_categorical_accuracy: 0.7667
Epoch 22/500
101/101 [=====] - 0s 2ms/step - loss: 0.6867 - categorical_accuracy: 0.7602 - val_loss: 0.6807 - val_categorical_accuracy: 0.7717
Epoch 23/500
101/101 [=====] - 0s 2ms/step - loss: 0.6797 - categorical_accuracy: 0.7627 - val_loss: 0.6737 - val_categorical_accuracy: 0.7742
Epoch 24/500
101/101 [=====] - 0s 2ms/step - loss: 0.6735 - categorical_accuracy: 0.7648 - val_loss: 0.6671 - val_categorical_accuracy: 0.7835
Epoch 25/500
101/101 [=====] - 0s 2ms/step - loss: 0.6678 - categorical_accuracy: 0.7665 - val_loss: 0.6609 - val_categorical_accuracy: 0.7885
Epoch 26/500
101/101 [=====] - 0s 2ms/step - loss: 0.6625 - categorical_accuracy: 0.7689 - val_loss: 0.6552 - val_categorical_accuracy: 0.7897
Epoch 27/500
101/101 [=====] - 0s 2ms/step - loss: 0.6577 - categorical_accuracy: 0.7687 - val_loss: 0.6504 - val_categorical_accuracy: 0.7891
Epoch 28/500
101/101 [=====] - 0s 2ms/step - loss: 0.6533 - categorical_accuracy: 0.7711 - val_loss: 0.6463 - val_categorical_accuracy: 0.7885
Epoch 29/500
101/101 [=====] - 0s 2ms/step - loss: 0.6490 - categorical_accuracy: 0.7739 - val_loss: 0.6425 - val_categorical_accuracy: 0.7860
Epoch 30/500
101/101 [=====] - 0s 2ms/step - loss: 0.6455 - categorical_accuracy: 0.7747 - val_loss: 0.6381 - val_categorical_accuracy: 0.7922
Epoch 31/500
101/101 [=====] - 0s 2ms/step - loss: 0.6421 - categorical_accuracy: 0.7758 - val_loss: 0.6344 - val_categorical_accuracy: 0.7971
Epoch 32/500
101/101 [=====] - 0s 2ms/step - loss: 0.6385 - categorical_accuracy: 0.7782 - val_loss: 0.6309 - val_categorical_accuracy: 0.7959
Epoch 33/500
101/101 [=====] - 0s 2ms/step - loss: 0.6353 - categorical_accuracy: 0.7803 - val_loss: 0.6280 - val_categorical_accuracy: 0.7953
Epoch 34/500
101/101 [=====] - 0s 2ms/step - loss: 0.6325 - categorical_accuracy: 0.7810 - val_loss: 0.6251 - val_categorical_accuracy: 0.7990
Epoch 35/500
101/101 [=====] - 0s 2ms/step - loss: 0.6297 - categorical_accuracy: 0.7841 - val_loss: 0.6238 - val_categorical_accuracy: 0.8058
Epoch 36/500
101/101 [=====] - 0s 2ms/step - loss: 0.6270 - categorical_accuracy: 0.7823 - val_loss: 0.6203 - val_categorical_accuracy: 0.7978
Epoch 37/500
101/101 [=====] - 0s 2ms/step - loss: 0.6248 - categorical_accuracy: 0.7851 - val_loss: 0.6180 - val_categorical_accuracy: 0.7984
Epoch 38/500
101/101 [=====] - 0s 2ms/step - loss: 0.6220 - categorical_accuracy: 0.7868 - val_loss: 0.6151 - val_categorical_accuracy: 0.8021
Epoch 39/500
101/101 [=====] - 0s 2ms/step - loss: 0.6199 - categorical_accuracy: 0.7861 - val_loss: 0.6138 - val_categorical_accuracy: 0.8108
Epoch 40/500
101/101 [=====] - 0s 2ms/step - loss: 0.6178 - categorical_accuracy: 0.7875 - val_loss: 0.6112 - val_categorical_accuracy: 0.8033
Epoch 41/500
101/101 [=====] - 0s 2ms/step - loss: 0.6157 - categorical_accuracy: 0.7883 - val_loss: 0.6087 - val_categorical_accuracy: 0.8071
Epoch 42/500
101/101 [=====] - 0s 2ms/step - loss: 0.6136 - categorical_accuracy: 0.7893 - val_loss: 0.6065 - val_categorical_accuracy: 0.8096
Epoch 43/500
101/101 [=====] - 0s 2ms/step - loss: 0.6116 - categorical_accuracy: 0.7898 - val_loss: 0.6048 - val_categorical_accuracy: 0.8114
Epoch 44/500
101/101 [=====] - 0s 2ms/step - loss: 0.6102 - categorical_accuracy: 0.7916 - val_loss: 0.6037 - val_categorical_accuracy: 0.8089
Epoch 45/500
101/101 [=====] - 0s 2ms/step - loss: 0.6080 - categorical_accuracy: 0.7925 - val_loss: 0.6017 - val_categorical_accuracy: 0.8133
Epoch 46/500
```


[illegible]

[illegible]

[illegible]

19

[illegible]

21

22

[illegible]

24