

Memoria Práctica 3

Jorge Benjumea
Ibai Burgos
Miguel Ángel Serrano

7 de Enero de 2024
Universidad Politécnica de Madrid

Abstract

Práctica sobre redes neuronales

Contents

1	Autores y trabajo realizado	3
2	Experimentación	4
2.1	Aprendizaje	4
2.2	Resultados	5
2.3	Conclusiones	6

1 Autores y trabajo realizado

Nº Matrícula	Nombre	Apellidos	Porcentaje dedicación	Trabajo realizado
bt0089	Ibai	Burgos Sánchez	1/3	Todo
bt0354	Jorge	Benjumea Burillo	1/3	Todo
bt0094	Miguel Ángel	Serrano Sánchez	1/3	Todo

2 Experimentación

2.1 Aprendizaje

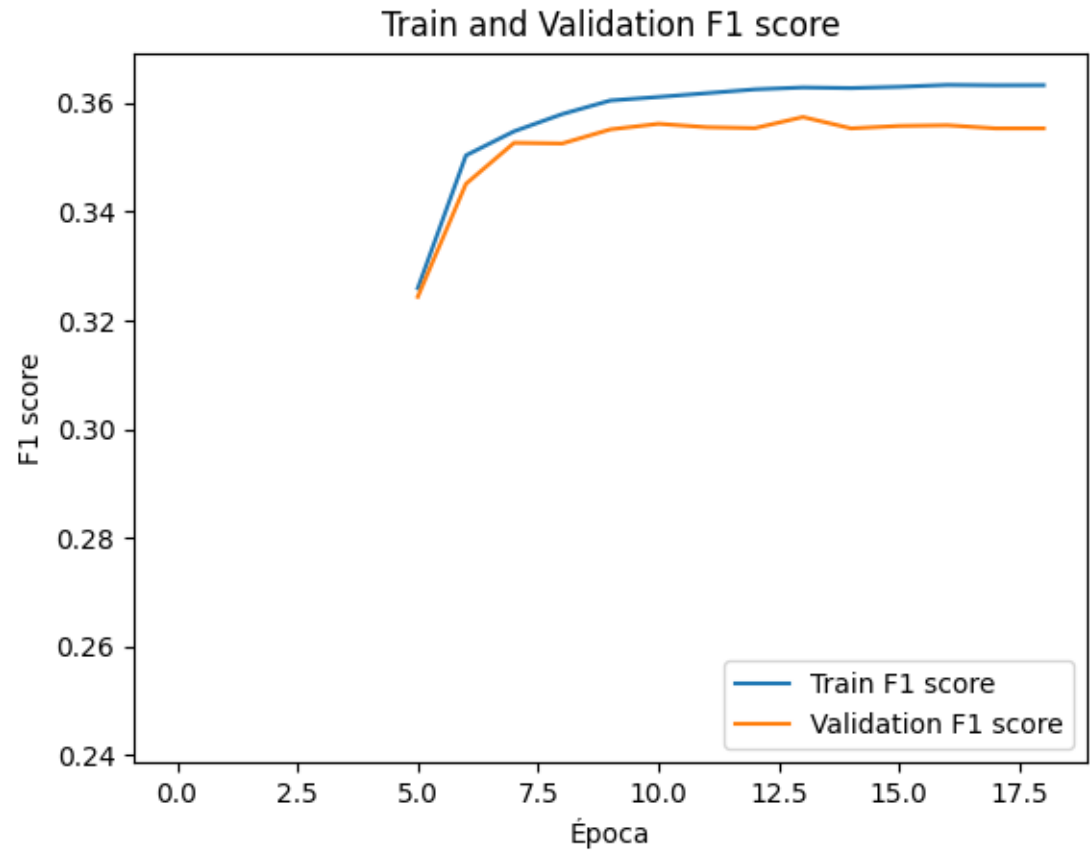


Figure 1: Ejemplo con 1 capa oculta, 5 neuronas y 0.1 de razón de aprendizaje

2.2 Resultados

- Todos los experimentos se han realizado con un seed = 10
- Todas las celdas que no tengan valor es porque no se ha conseguido que terminen (DNF)
- Las celdas de color rojo han sido ejecutadas con un Batch Size = 128

Table 1: Número de capas ocultas = 1

Número de Neuronas	5 Neuronas	10 Neuronas	20 Neuronas	100 Neuronas
Razón de Aprendizaje				
1	0.372011251758	0.373108060542	0.244550550134	0.244550550134
0.1	0.304868316041	0.374607603767	0.244550550134	0.244550550134
0.01	0.237588652482	0.251845419018	0.244550550134	0.244550550134
0.001	0.191768007483	0.244550550134	0.244550550134	0.244550550134

Cuando la capa oculta tiene 20 o más neuronas, parece que la red es incapaz de aprender, ya que siempre obtiene el mismo F1 score en los 5 ciclos. Además, no importa la razón de aprendizaje, siempre obtiene el mismo F1 score. También cabe indicar que, al contrario que con 5 o 10 neuronas, después de la primera pasada de entrenamiento, su resultado puede ser positivo en diabetes. Con las otras opciones, se necesitan varias pasadas de test hasta que sale como predicción diabetes, por lo que no se puede calcular la F1 score inicial ya que *precision* realiza una división por 0.

También se puede observar que una razón de aprendizaje menor o igual a 0.001 perjudica el rendimiento de la red. Esto se puede deber a que se quede en un mínimo local.

Según los experimentos, el mejor resultado se consigue con 5 o 10 neuronas y razón de aprendizaje 1. Se podría mejorar estos resultados guardando el estado de la red cada vez que mejora el F1 score, ya que se ha observado como a veces la red pierde mucho rendimiento tras un paso de entrenamiento. Además, es posible que esta configuración se beneficie cambiando la condición de parada a 5 ciclos sin mejorar el F1 score anterior, en vez del mejor.

Table 2: Número de capas ocultas = 3

Número de Neuronas	5 Neuronas	10 Neuronas	20 Neuronas	100 Neuronas
Razón de Aprendizaje				
1			0.2445505501349	0.244550550134
0.1			0.2445505501349	0.244550550134
0.01			0.2445505501349	0.244550550134
0.001			0.2445505501349	0.244550550134

No ha sido posible entrenar la red con 5 o 10 neuronas. Además, por alguna razón, la red parece no aprender y además da el mismo resultado que al utilizar una sola capa oculta y 20 o 100 neuronas. No completamos el resto de las celdas de 20 y 100 neuronas porque con los resultados obtenidos el valor parece repetirse y tarda en completar el entrenamiento.

Table 3: Número de capas ocultas= 10

Número de Neuronas	5 Neuronas	10 Neuronas	20 Neuronas	100 Neuronas
Razón de Aprendizaje				
1			0.244550550134	0.244550550134
0.1			0.244550550134	0.244550550134
0.01			0.244550550134	0.244550550134
0.001			0.244550550134	0.244550550134

Mismas observaciones que en la tabla anterior

2.3 Conclusiones

Nos ha sido imposible conseguir entrenar a la red con 3 o 10 capas ocultas y 5 o 10 neuronas por capa. En cambio, si hemos conseguido entrenar a la red con 1 o 2 capas ocultas, por lo que pensamos que no se trata de un fallo en el código. Las posibles causas pueden ser no haber tratado lo suficiente los datos, que solo están normalizados y aleatorizados, o no haber dejado suficiente tiempo a la red.

Para entrenar a una red de 2 capas ocultas y 5 neuronas por capa, se necesitaron unas 3 horas como mínimo. Para 3 capas ocultas, el código estuvo ejecutándose entre 5 y 8 horas, sin predecir nunca diabetes, por lo que no se puede calcular la F1 score. Para mejorar el rendimiento, se hizo una prueba donde se ejecutaban pasadas de entrenamiento con 128 muestras simultáneas, sin modificar la red, y luego se modifica al final de las 128 muestras. Con este cambio, se consiguió entrenar la red de 2 capas ocultas en 2 minutos. Sin embargo, la de 3 capas estuvo 2 horas y media sin éxito.

Para intentar que la F1 score se disponía de lo antes posible, se redujeron los datos de no diabetes a la misma cantidad que los con diabetes, ya que, al haber tanta diferencia, puede que la red tiende a aprender primero por ser más. La selección de los casos de no diabetes era al azar y por cada época, para

intentar que la red entrenase con todos los ejemplos. Sin embargo, este cambio fue perjudicial, ya que hasta una vez que obtuvo una F1 score válido, volvió a obtener una inválida después, algo que nunca había pasado antes. Lo más sencillo hubiese sido invertir la F1 score y hacerlo con los no diabéticos, pero en teoría lo hacen solo con los positivos, por lo que no estábamos seguros si era válido. Tampoco tiene gran importancia, ya que un F1 score inválido no para la ejecución antes, ya que no lo contamos, y la red utiliza la función *loss_function* para el aprendizaje.

Con los datos obtenidos, pensamos que lo mejor es utilizar una red con 1 capa oculta y 10 neuronas por capa y razón de aprendizaje de 1, implementando los cambios propuestos a analizar anteriormente como guardar la red cada vez que mejora y solo pararla si lleva 5 ciclos sin mejorar su resultado anterior.

Cabe destacar que, con el cambio de realizar la pasada de entrenamiento con 128 muestras simultáneamente, se mejora mucho el rendimiento y no parece afectar a la F1 score, por lo que recomendamos usarlo.

[PÁGINA DEJADA INTENCIONALMENTE EN BLANCO]