

Introducción a Redes Neuronales - Proyecto

David Segura, #13-11341

Julio 17, 2020

1 Introducción

Las redes neuronales artificiales son un modelo que emulan la forma en la que el cerebro humano funciona mediante combinaciones de parámetros y aplicación de las mismas, donde a través de cierto entrenamiento se intenta encontrar la mejor combinación posible, siendo esto el proceso de aprendizaje. Cuando una red ya ha sido entrenada, esta se puede usar para realizar predicciones o clasificaciones de datos, entre muchas aplicaciones más, pero en esta ocasión son las que trataremos para resolver el problema presentado.

Hemos seleccionado los datos *Default of Credit Card Clients*[1] correspondientes a 30.000 tarjetas de crédito que debemos clasificar. Para ello usaremos las distintas técnicas aprendidas en el curso de *CO-6612 Introducción a las Redes Neuronales*, dictada por la profesora Minaya Villasana.

2 Procesamiento de Datos

2.1 Análisis de los Datos

Al estudiar los datos pudimos observar que cada patrón está compuesto de 23 atributos, en los que se proporciona información bancaria de un cliente, entre los cuales tenemos campos como: *balance límite*, *edad*, *estado civil*, *educación*, *sexo*, *cuentas de pago* y también se nos proporciona el valor deseado. Todos estos valores representados numéricamente.

2.2 Escalamiento

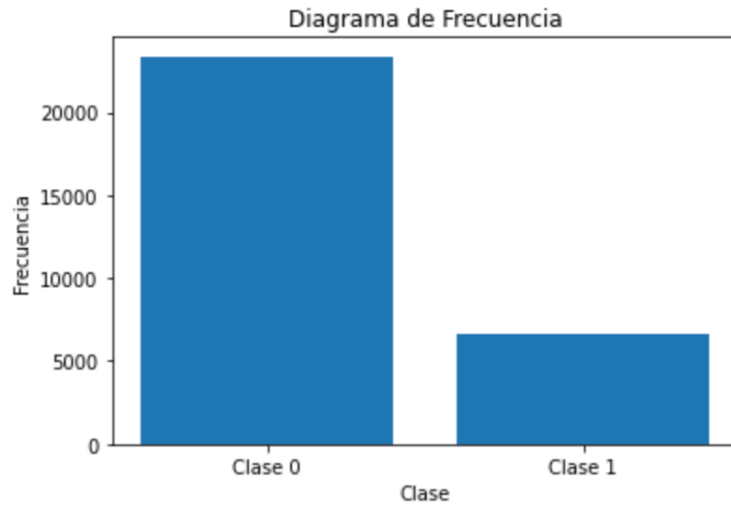
Se decidió realizar un escalamiento de los datos debido a que la variación entre los números es muy alta, entonces se cada número en un rango de $[0,1]$, siguiendo la siguiente fórmula:

$$x^{(k)} = \frac{x_{max}^{(k)} - x^{(k)}}{x_{max}^{(k)} - x_{min}^{(k)}}$$

donde $x^{(k)}$ es un atributo perteneciente a la k -ésima columna de nuestra matriz con todos los patrones, siendo $x_{max}^{(k)}$ el atributo de mayor valor de la columna k y $x_{min}^{(k)}$ el menor.

2.3 Clases

En los valores deseados se extrajo que los patrones son pertenecientes a una de 2 clases: 0 ó 1. Consideraremos la cantidad de elementos que se nos proporciona por cada clase:



La *Clase 0* contiene 23.364 datos, mientras la *Clase 1* tiene 6.636 datos. Esta proporcionalidad que resalta se considera que puede dificultar el aprendizaje de las redes debido a la *poca* información proporcionada por la *Clase 1*.

2.4 Particionamiento

Para la mayoría de las clasificaciones, tomamos la decisión de particionar los datos tomando $\frac{2}{3}$ de ellos para el entrenamiento y $\frac{1}{3}$ para los datos de validación, asegurandonos de que para ambos casos queden datos pertenecientes de cada clase.

Al final nos quedan 20.000 patrones para usar en el entrenamiento y 10.000 que nos quedan como datos de validación.

3 MLP

El *MultiLayer Perceptron* fue la primera red que decidimos usar para realizar el entrenamiento de clasificación de datos, usando una arquitectura de 4 capas, 1 de input, 2 capas ocultas de 100 y 50 neuronas respectivamente, con ambas funciones de activación siendo la *ReLU* y 1 capa de salida de una neurona con función de activación *sigmoid*, usando el optimizador *Adadelta* con una tasa de aprendizaje de 0.1. Con estas configuraciones encontramos la mejor precisión con los datos de validación y fue de 81.29%.

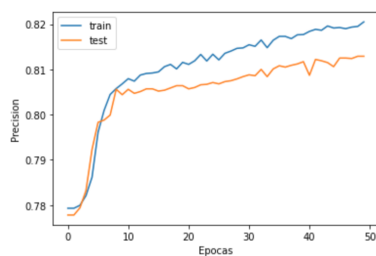


Figure (2) Accuracy

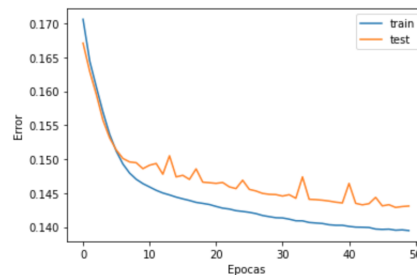


Figure (3) EMC

También se probaron otros optimizadores como Descenso del Gradiente y Adam, pero se consiguieron peores resultados.

4 Boosting con MLP

Se usó la estrategia proporcionada por la profesora Minaya Villasana en sus láminas[2]. Para ello se realizó una nueva partición de los datos, en este caso 4 particiones, uno para entrenar a cada uno de los expertos $E1$, $E2$ y $E3$, y otra cuarta partición para los datos de validación, quedando la distribución de 8.571, 8.571, 8.572 y 4.286 datos respectivamente. La arquitectura escogida para cada uno de los expertos es la misma a la usada en el MLP con las mismas configuraciones.

La precisión conseguida con esta estrategia fue de 82.12%, apenas 0.83% mejor en cuanto a la encontrada con MLP.

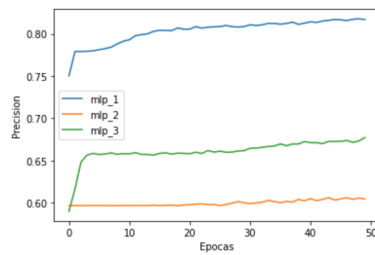


Figure (4) Accuracy Training

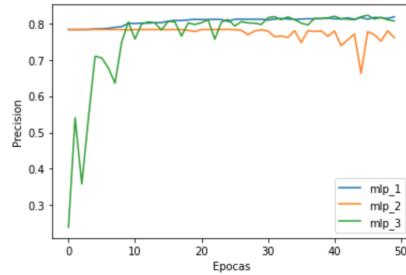


Figure (5) Accuracy Test

5 SVM

La última estrategia que usamos es *Support Vector Machine*, con regularización 1 y kernel Gaussiano consiguiendo una precisión con los datos de validación de 82.02%.

También clasificamos con el kernel polinómico de grado 2 y obtuvimos un mejor resultado que fue de 82.26%.

6 Conclusión

Después de probar las estrategias usadas durante el desarrollo de la misma, el mejor clasificador que se pudo encontrar fue el SVM con kernel polinómico de grado 2, alcanzando una precisión del 82.26%.

Considero que los datos tuvieron un papel importante en el entrenamiento de las redes, debido a la desproporcionalidad mencionada en el número de datos correspondiente a cada clase. Se contó con una cantidad de datos de la clase 1 aproximadamente 3 veces menor a la cantidad de datos de la clase 0, haciendo la tarea difícil de las redes de aprender a clasificar los datos.

References

- [1] Yeh, I. C., & Lien, C. H. (2009). The comparisons of data mining techniques for the predictive accuracy of probability of Default of credit card clients. *Expert Systems with Applications*, 36(2), 2473-2480.
- [2] Villasana, Minaya (2020). Perceptrón Multicapas (Backpropagation). Lámina 81.