

Checkpoint 4 - Grupo 33

Introducción

Para empezar le agregamos una semilla al sistema de keras para luego poder replicar resultados. Luego, construimos un modelo muy simple con pocas neuronas, batches y épocas, con regularización L2 y ningún tipo de optimización ni dropout, pero con 5 capas ocultas. Este resultado fue bastante alto así.

A partir de esto hicimos Grid Search para encontrar las mejores combinaciones entre batches, épocas, función de activación para las capas ocultas y de esta forma conseguir la mejor optimización del modelo. Este programa nos recomendaba siempre parámetros altos para batches y épocas.

También intentamos utilizar L1 pero nos dimos cuenta que su impacto sobre L2 era mínimo. Además, experimentamos con el Learning Rate pero nos dimos cuenta que un valor mayor o igual a 0.01 ya hacía que el modelo pegue muchos saltos, esto se podía apreciar en el "Loss" que iba subiendo y bajando erráticamente con cada época, en el diagrama de "Época vs Loss".

El mayor impacto para el modelo fue la cantidad de neuronas que elegimos para entrenar. Por mucho tiempo experimentamos con 120 neuronas por capa, pero con este número nos saltaban varios Predicted False por lo cual tuvimos que cambiar el umbral de 0.5 a 0.4 para balancear los resultados. Igualmente, más adelante nos dimos cuenta que el modelo llegó a sobreentrenar con los datos de train, por lo cual era mejor no solo bajar la cantidad de neuronas por capa, sino que además bajar el número de capas ocultas a 3, esto último tomó consecuencias muy positivas en nuestro modelo, logrando con esto la mejor predicción de Kaggle.

Construcción del modelo

Nuestro mejor modelo fue construido de la siguiente manera:

Tenemos una capa de entrada con 60 neuronas, sin función de activación por lo cual se comporta de manera lineal. Esta capa, al no tener Dropout es Fully Connected.

Luego tenemos 3 capas ocultas, cada una también con 60 neuronas, con función de activación “ReLU” y con un DropOut de 0.3.

Finalmente la capa de salida tiene solamente una neurona ya que trabajamos con una salida binaria y con función de activación “Sigmoide” ya que es la más ideal para clasificaciones binarias.

Todas las capas utilizan el método de regularización L2 y como inicializador de pesos “Glorot Normal”. Utilizamos este inicializador para setear los pesos iniciales de las conexiones que investigamos que es el mejor para ya que toma en cuenta las entradas.

Con Grid Search buscamos los mejores valores hiperparametros como batches, épocas, optimizador y función de activación para las capas ocultas. Los mejores valores para épocas y batches fueron 60 y 50 respectivamente. El mejor optimizador resultó ser ‘NAdam’ y la funcion de activacion fue “ReLU”.

Como último detalle, en la etapa de compilación utilizamos la función de Loss “Binary Crossentropy” ya que estamos clasificando valores binarios.

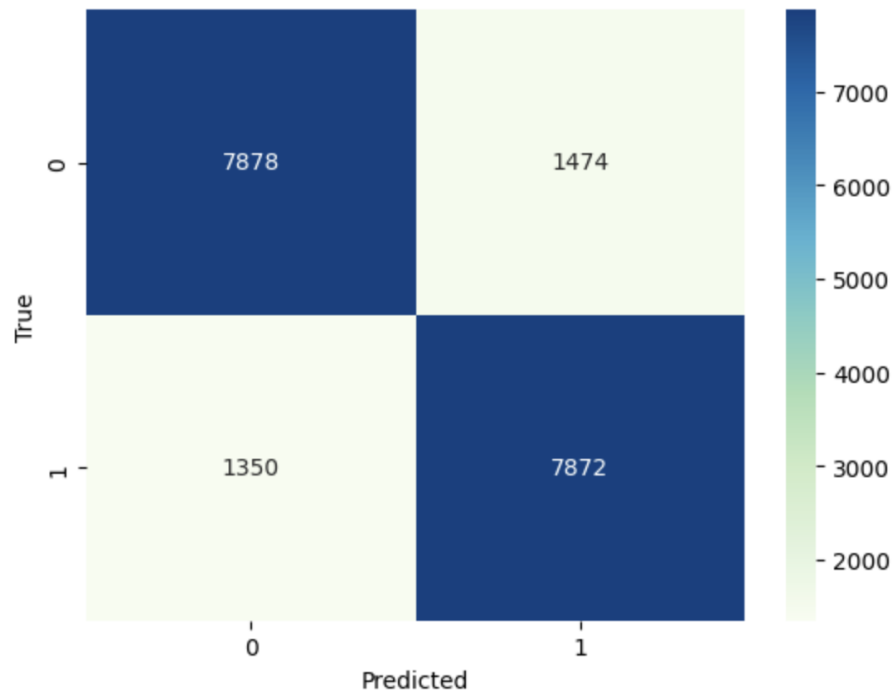
Cuadro de Resultados

| Modelo | F1-Test | Precision Test | Recall Test | Accuracy | Kaggle |
|----------|------------------------|------------------------|------------------------|------------------------|---------|
| modelo_1 | 0.84811849915 59112 | 0.85187616234 54765 | 0.84439384 08154413 | 0.84984386 77721546 | 0.8501 |
| modelo_2 | 0.84791038345 54071 | 0.8422854697 196661 | 0.85361093 03838647 | 0.847959513 2981587 | 0.84899 |
| modelo_3 | 0.8404804608 946139 | 0.8471567875 53413 | 0.83390854 18464194 | 0.841983417 6806288 | 0.84129 |
| modelo_4 | 0.8494275886 852507 | 0.8102046812 261287 | 0.89264134 6568839 | 0.84209109 50791429 | 0.8401 |

El mejor modelo terminó siendo el que tiene menor cantidad de neuronas por capa y cantidad de capas ocultas. Los otros son modelos con más complejidad ya que además de más capas, batches y épocas, también usábamos cross validation. Toda esta

combinación de decisiones terminaron creando modelos que estaban sobre-entrenando y por más que las métricas en el set de validación eran buenas al subirlas a Kaggle disminuían, el tiempo que tardaron en entrenarse no era demasiado extenso ya que utilizaban una learning rate mayor.

Matriz de Confusion



La cantidad de falsas predicciones está bastante balanceada por lo cual decidimos dejar el umbral en 0.5 en vez de tratar de compensar.

Tareas Realizadas

| Integrante | Tarea |
|-----------------------|--------------------------------------|
| Lucas Nahuel Raimondi | Modelo Base + Optimización + Informe |
| Dolores Levi | Optimización + Informe |
| Manuel Davila | Optimización + Informe |