

Checkpoint 4 - Grupo 16

Introducción

Para la realización del presente trabajo, se tomó como base el dataset depurado obtenido en el “checkpoint1”, el cual fue normalizado con una distribución normal estándar. Se entrenaron 3 modelos: un perceptrón simple, un perceptrón simple y una red neuronal multicapa que se comparó el rendimiento con y sin early stopping.

Construcción del modelo

Para el mejor modelo obtenido, llamado en el cuadro de abajo **“Perceptrón multicapa”**, normalizamos los datos con una distribución normal estándar y elegimos la siguiente arquitectura:

- *Capa de entrada:* una capa de estilo *InputLayer*, que sólo toma las variables de entrada y las pasa a la capa oculta, tiene el mismo tamaño que la cantidad de columnas luego de preprocesar los datos (209 entradas). Este objeto no utiliza una función de activación, se puede entender que “deja pasar las variables a la capa siguiente”. Está fully-connected con la capa oculta.
- *Capas ocultas:* se utilizó una sola capa oculta, fully-connected con la capa de input y la capa de output. Se decidió mantener la misma cantidad de neuronas que de variables de entrada (209) y se le dió a esta la función de activación ReLU (Rectified Linear Unit).
- *Capa de salida:* consta de una neurona que tiene una función de activación sigmoidea para la realización de la predicción de probabilidad de un resultado binario (cancela o no cancela). Luego, a estas probabilidades se las interpreta como que un porcentaje mayor a 50% es predecir un éxito (un 1).

La función de loss elegida para el modelo fue binary cross-entropy y para minimizarla se optó por el optimizador Adam con un learning rate de 0.00001. El modelo pasó 500 épocas intentando minimizar el loss tomando los datos de a “tandas” (batches) de 32. La red neuronal entrenó con un 80% del total del split de train, por lo que se tenía un 20% de datos para vigilar el overfitting durante el correr de las epochs.

En contraste con los modelos **“RNA con early stopping”** y **“RNA sin early stopping”**, para este modelo no se utilizó ninguna técnica de regularización ni tampoco se optimizaron hiperparámetros, teniendo esto en cuenta y la arquitectura más simple, este modelo fue el que performó mejor de todas maneras.

Cuadro de Resultados

Presentamos un cuadro de resultados donde exponemos las métricas F1, Precision y Recall que obtuvimos de nuestros modelos en test y la métrica F1 obtenida de las predicciones subidas a Kaggle.

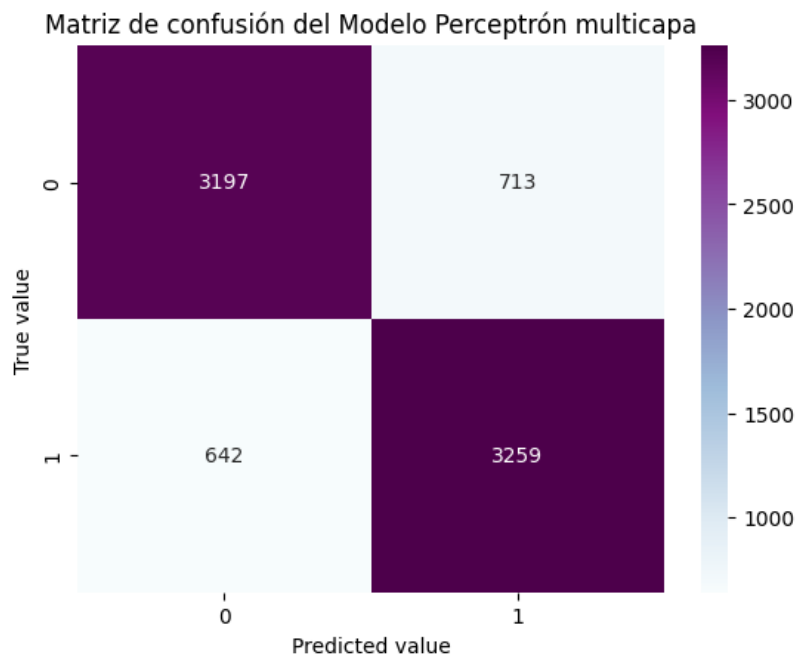
Modelo	F1-Test	Presicion Test	Recall Test	Kaggle
Perceptrón simple	0.7810	0.7794	0.7826	0.79044
Perceptrón multicapa	0.8278	0.8204	0.8354	0.82457
RNA con early stopping	0.8273	0.8184	0.8364	0.8226
RNA sin early stopping	0.8313	0.8182	0.8449	0.81457

Descripción de modelos:

- Perceptrón simple: red neuronal con una capa de entrada InputLayer fully-connected de 209 inputs y luego directamente una capa de salida de una neurona con función de activación sigmoidea (no tiene capas ocultas). Usa el optimizador Adam y fue entrada por 50 epochs con un batch size de 32.
- Perceptrón multicapa: red neuronal con una capa de entrada InputLayer fully-connected de 209 inputs, una capa oculta de 209 inputs con activación ReLU también fully-connected y luego una capa de salida de una neurona con función de activación sigmoidea. Usa el optimizador Adam y fue entrada por 500 epochs con un batch size de 32.
- RNA con early stopping: red neuronal con una capa de entrada InputLayer fully-connected de 209 inputs, 4 capas ocultas de cantidad de neuronas 189, 164, 203 y 68 secuencialmente, con activación LeakyReLU de parámetro 0.05, dropout y regularización L2, luego una capa de salida de una neurona con función de activación sigmoidea. Usa el optimizador Nadam y fue entrada por 27 epochs con un batch size de 29, este se detuvo por el early stopping de paciencia 8. Cantidad de neuronas, rates de dropout, learning rate, batch size y coeficientes de regularización L2 fueron optimizados de manera bayesiana (de la misma manera que el checkpoint 3) con el objetivo de maximizar el f1 score.
- RNA sin early stopping: red neuronal con la misma arquitectura e hiperparámetros que “RNA con early stopping” pero esta versión no tiene early stopping y se entrenó durante 100 epochs.

Matriz de Confusión

La matriz de confusión para el mejor modelo, que fue el **Perceptrón multicapa** es la siguiente:



Tareas Realizadas

Integrante	Tarea
DIEM, Walter Gabriel	Perceptrón simple, Perceptrón multicapa, RNA con early stopping, RNA sin early stopping, Informe
MAIOLO, Alejandro	Perceptrón multicapa, Informe
RUIZ, Karen Belén	Perceptrón simple, Informe