

Grupo 5

Santiago Cifuentes, Felipe Torres, Andres Hernandez, Juan Carlos Franco

Para el challenge implementamos dos métodos de clasificación. El primer modelo entrenado fue un clasificador de tipo random forest con 14 estimadores, máximo 14 características y 20 muestras por hoja; así mismo el modelo se validó por medio de K-fold validation utilizando 5 folds y aleatorizando los datos. El segundo clasificador implementado fue una red completamente conexa con 10 capas ocultas usando 8 neuronas en cada una y una función de activación de tipo sigmoide; entrenando con Adam optimizer, un learning rate inicial de $1e-3$, durante 500 épocas. Para probar el modelo, los datos suministrados se dividieron en un 90% para entrenamiento y 10% para su validación (90/10). El método con mayor precisión resultó ser el de random forest con $\approx 97\%$ de clasificación correcta. Para probar la sensibilidad del modelo respecto al número de datos usados en entrenamiento se cambiaron las fracciones previamente mencionadas de 90/10 a 60/40 y se encuentra una pérdida no mayor a 1% en la clasificación correcta. La tabla a continuación presenta los resultados obtenidos:

Tabla 1: Resultados obtenidos al entrenamiento de 5 modelos. Se mide la precisión promedio de clasificación para elegir el mejor modelo.

Random State	Random Forest	K-fold Validation	Redes
2	0,963	0,962	0,965
10	0,968	0,962	0,960
22	0,969	0,961	0,968
124	0,974	0,962	0,966
200	0,969	0,964	0,958
1000	0,962	0,961	0,955
Promedio	0,967	0,962	0,962