



Facultad de Ciencias Exactas, Ingeniería y Agrimensura
UNIVERSIDAD NACIONAL DE ROSARIO

Introducción a la Inteligencia Artificial

Trabajo Práctico Aprendizaje Automatizado

Alumnas:

Cipullo, Inés

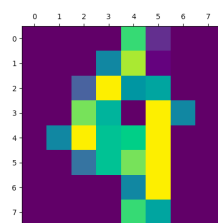
Sullivan, Katherine

2022

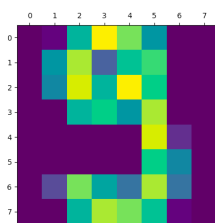
Ejercicio 1

El conjunto de datos a utilizar es de imágenes de los dígitos del 0 al 9. Cada dato es una imagen de 8x8 y se representa mediante una lista de 64 píxeles. El objetivo es realizar un aprendizaje supervisado sobre dicho conjunto, buscando clasificar cada dato en la clase del respectivo dígito que muestra.

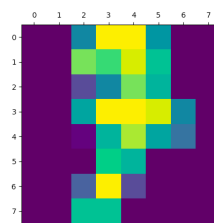
El conjunto de datos contiene 1797 muestras en total y habiendo 10 clases, se cuenta con aproximadamente 180 muestras por clase. Las siguientes son muestras etiquetadas de los datos del conjunto.



(a) Etiqueta: 4



(b) Etiqueta: 9



(c) Etiqueta: 7

Ejercicio 2

a.

Al entrenar un árbol de decisión con los parámetros por defecto se obtiene en todos los casos un valor de accuracy sobre los datos de entrenamiento igual a 1, y un valor de accuracy sobre los datos de evaluación que ronda el 0.86, en ningún caso de los testeados superando el 0.9.

Si bien uno podría pensar que un accuracy de 0.86 es más que aceptable, la

diferencia notable con respecto a la accuracy sobre el conjunto de entrenamiento (más teniendo en cuenta que esta es perfecta), es un claro indicador de sobreentrenamiento.

Con el objetivo de buscar clasificar cada dato de entrenamiento sin equivocaciones, un modelo puede terminar siendo más complejo de lo que la generalización debe ser. Esto es lo que se puede observar al graficar el árbol obtenido, que resulta de grandes dimensiones, con muchas de sus hojas clasificando solo un elemento del conjunto de datos.

Este comportamiento se da porque los parámetros de parada por defecto presentan restricciones laxas o nulas. Podemos visualizar esto con, por ejemplo, el parámetro `max_depth` que establece un cota a la cantidad de niveles del árbol (profundidad) cuyo valor por defecto es `None`, o el parámetro `min_samples_leaf` que es una restricción sobre la cantidad mínima de elementos que debe clasificar una hoja del árbol y su valor por defecto es `1`.

b.

Se buscaba realizar un análisis de como varían los resultados de un modelo dependiendo de los parámetros de parada con los que se define. Para ello primero definimos qué parámetros de parada tendremos en cuenta. Estos son:

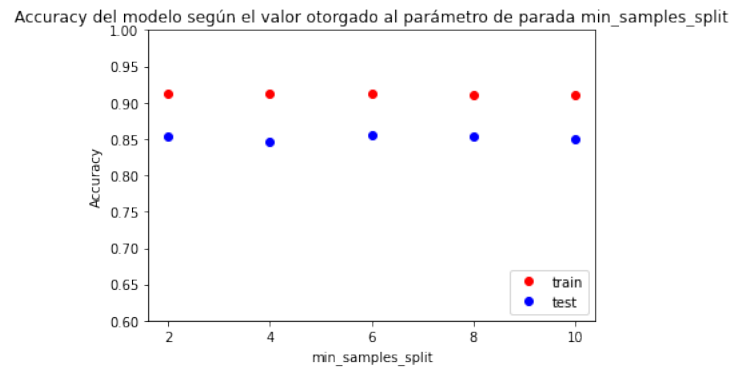
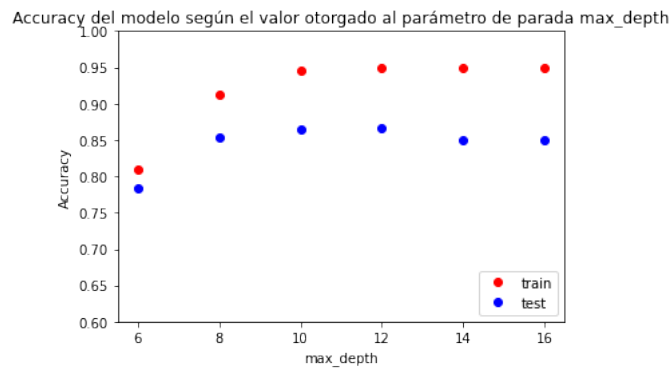
- `max_depth`
- `min_samples_split`
- `min_samples_leaf`
- `max_leaf_nodes`

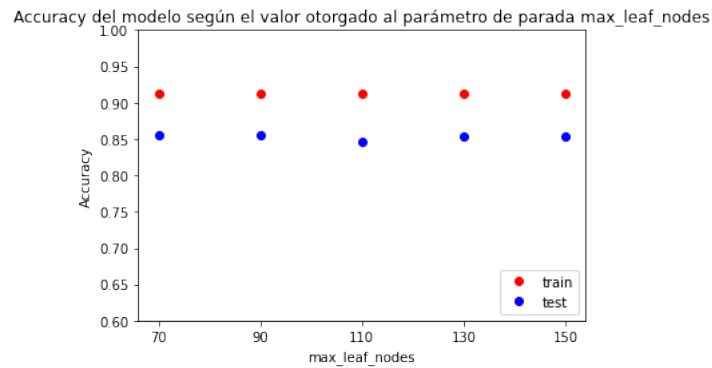
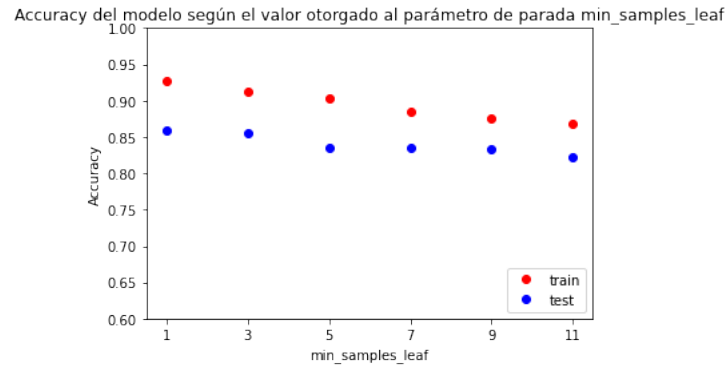
Luego se entrenaron distintos modelos haciendo variar estos parámetros en rangos de valores significativos en cada caso y se evaluó cuales son los valores óptimos de cada parámetro. Si bien no hay tendencias claras de que haya un

único valor óptimo para cada parámetro dado que estos varían, una muy buena base de valores resulta:

- `max_depth = 8`
- `min_samples_split = 4`
- `min_samples_leaf = 3`
- `max_leaf_nodes = 110`

Para construir gráficas significativas, utilizaremos estos valores, manteniendo 3 de ellos fijos y haciendo variar el restante. De ahí surgen los siguientes gráficos:



**C.**

Al entrenar un modelo con sus parámetros por defecto no es posible lograr una accuracy que supere el 80% sin sufrir sobreentrenamiento, entendiéndose con sobreentrenamiento a que la diferencia entre el accuracy sobre el conjunto de entrenamiento y evaluación sea menor al 5%. Luego, al hacer variar los cuatro

parámetros de parada considerados en rangos significativos para cada caso, vemos que si es posible lograr 0.8 o más de accuracy sin sobreentrenamiento pero no siempre.

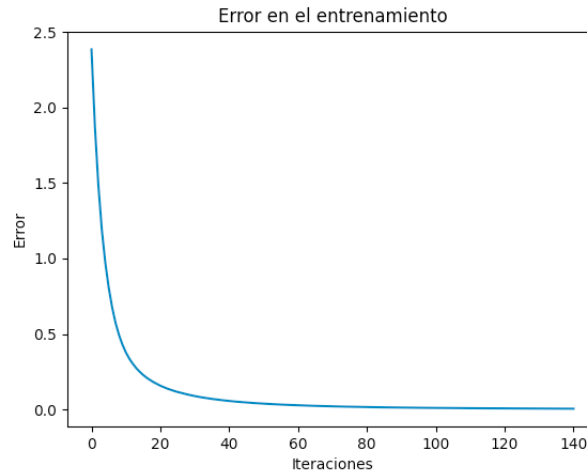
Si bien, los árboles de decisión son modelos propensos a sufrir de sobreentrenamiento, es un comportamiento que puede ser controlado. Para este conjunto de datos, con los parámetros de parada adecuados se puede evitar dicha conducta, lo cual resulta algo esperado, pero normal que sea extraño.

Ejercicio 3

a.

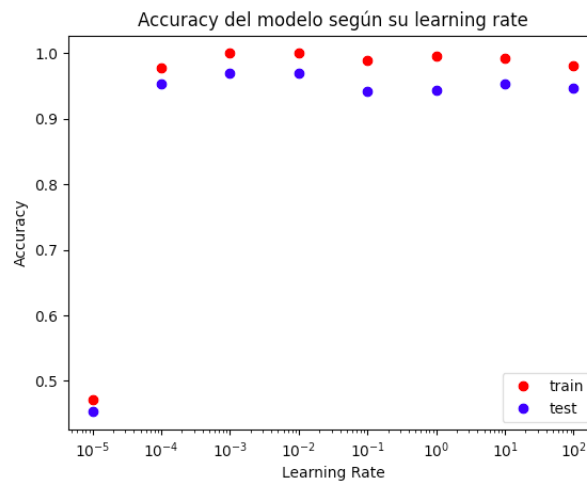
Al entrenar una red neuronal con los hiperparámetros por defecto se obtiene en todos los casos un valor de accuracy sobre los datos de entrenamiento igual a 1, y el valor de accuracy sobre los datos de evaluación siempre superior a 0.96 (en los casos testeados), con un promedio de casi 0.98.

Los resultado que se obtienen resultan más que aceptables, ya que si bien la accuracy sobre los datos de entrenamiento es perfecta, en este caso no compromete la accuracy sobre los datos de evaluación, que resulta también muy elevada. En el siguiente gráfico se muestra la curva que realiza el valor del error del modelo según la iteración del entrenamiento. Notamos que luego de unas pocas iteraciones el error del modelo resulta despreciable.



b.

Se presenta a continuación un gráfico que muestra la accuracy sobre el conjunto de datos de entrenamiento y la accuracy sobre el conjunto de datos de prueba, según el learning rate con el que se entrenó el modelo, variando este último entre los siguientes valores: 0.00001, 0.0001, 0.001, 0.01, 0.1, 1, 10.



El learning rate de un modelo determina el tamaño de los pasos de aprendizaje que se van dando entre iteraciones, es decir, que tanto afecta la actualización de los pesos de las conexiones neuronales al peso actual, ante un nuevo paso de aprendizaje. En general, el valor del learning rate es pequeño (menor a 1) para dar pasos chicos. Sin embargo, ambos extremos resultan perjudiciales al modelo. La tendencia es que valores muy bajos de learning rate hacen que el entrenamiento del modelo sea muy lento y hasta se pueda estancar en alguna medida intermedia, nunca llegando a valores de error aceptables, y valores muy elevados llevan rápidamente a sobreentrenamiento y los resultados pueden desestabilizarse.

En este caso particular, notamos que para los learning rates con los que se evaluó, los valores de accuracy resultan elevados, todos mayores a 0.9, excepto para el learning rate 0.00001, en el que los valores de accuracy están por debajo de 0.5. Los learning rates que mejores resultados ofrecieron fueron 0.001 y 0.01, coincidiendo en una accuracy perfecta sobre el conjunto de entrenamiento y una accuracy de 0.96944 sobre el conjunto de evaluación

Es posible visualizar las observaciones teóricas en el análisis realizado sobre el conjunto de datos con el que se trabajó.