



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

Introducción a la Inteligencia Artificial

Trabajo Práctico Aprendizaje Automatizado

Alumnas:

Cipullo, Inés

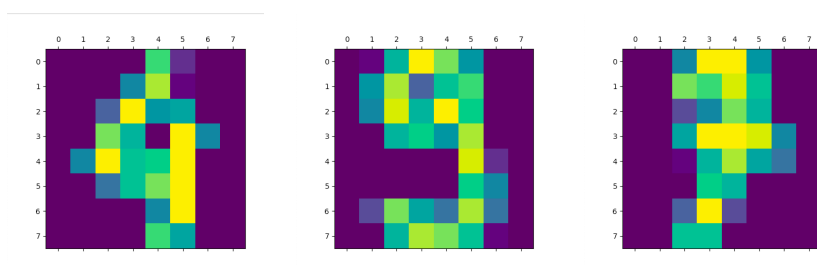
Sullivan, Katherine

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 el valor de accuracy sobre los datos de evaluación 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 y 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`.