Entrega parcial

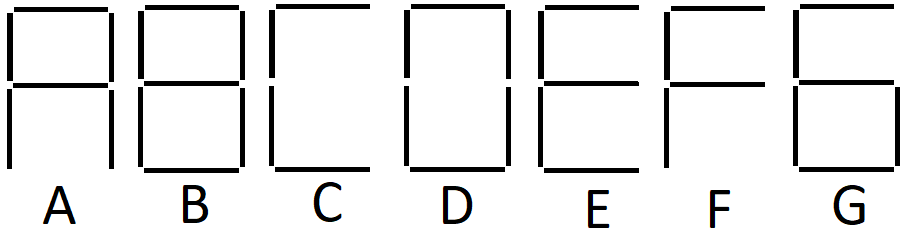
Vamos ha utilizar el dataset de “letritas”.

Para validar los modelos se ha pensado utilizar *Validación Cruzada*.

Se ha pensado crear datasets utilizando cuatro formas distintas de tratar los datos: Cuadrados, Patrones, Diferencia y N.º de pixeles con un valor inferior a 127.

Tratar las imágenes por Cuadrados, zonas o grupos consiste en crear los atributos del dataset en función de la mayoría de los pixeles blancos o negros de una zona. Las formas de crear las zonas o grupos de píxeles serían: columnas, filas, elección aleatoria (p. ej.: los píxeles 11, 125, 4566, 4587, 7236, son el atributo 1, …), grupos por cuadrados (p. ej.: del 0 al 5, del 100 al 105, del 200 al 205, del 300 al 305 y del 400 al 405 son el atributo 1, …) y diagonales.

Tratar las imágenes por Patrones consiste procesar las imágenes y pasarlas la codificación que haría una calculadora de donde se marcan las líneas de un 8, en función del número que sea. Al ser letras la codificación sería:



Cada lado sería un atributo, así que habría 7 atributos.

Tratar las imágenes por Diferencia consiste en comparar píxel a píxel o zonas de píxeles de las imágenes de letras dibujadas a mano con las letras que han sido dibujadas por un ordenador (la primera fila). Los atributos serían el resultado de la comparación, si son iguales, distintos o un porcentaje de parecido.

Tratar las imágenes por N.º de pixeles con un valor inferior a 127 consiste en contar el número de pixeles que cumplen la condición y ese sería el único atributo.

Los modelos de clasificación que hemos pensado utilizar son los siguientes (de Scikit-learn): KNN, Regresión Logística, GaussianNB y Árboles de decisión.