

Algoritmo K-Vecinos más cercanos (KNN)

Propósito: Algoritmo de Aprendizaje Máquina que puede ser usado para clasificación o regresión.

1.- Cargar la instancia de datos

Ejemplo de un registro:

$R1 = [C1, C2, C3, \dots Cn, \text{Etiqueta}]$

Donde:

Cx = Una característica del registro

Etiqueta = Clase a la que pertenece el registro

2. Generar/Establecer un registro “j” para ser clasificado

3.- Establecer el valor de K. Donde K puede ser cualquier valor entero entre 1 y el total de registros inclusive.

4.- Para cada registro “i” en la instancia de datos

4.1.- Calcular la distancia entre el registro “j” y el registro “i”

Nota. La métrica de similitud usada para el cálculo de la distancia depende el problema a tratar. La distancia más común, es la distan Euclidiana.

4.2.- Añadir la distancia obtenida junto al registro “i” a una estructura de datos

5.- Ordenar la estructura de datos de acuerdo con la métrica de similitud. De tal forma que se encuentre en el primer índice el registro más similar al registro “j” y en el último índice al registro menos similar.

Nota. En la mayoría de las métricas de similitud un registro es más similar a otro si cuenta con una menor distancia y viceversa.

6. Recuperar de la estructura de datos ordenada los primeros K registros

7. Si el problema es de regresión, se devolverá la media de las etiquetas de los K registros

8. Si el problema es de clasificación, se devolverá la **moda** de las etiquetas de los K registros