



Tema 8

aprendizaje analógico: K-nn

wopert: Teorema de No Free Lunch "NFL" (se necesita muchos algoritmos para aplicar el mejor en cada caso)

Def Analogía:

KAGGLE es el github de las bases de datos

Lo de filtros no entra (filtros colaborativos)

Nosotros tenemos dos patrones A y B con una serie de características

para predecir un nuevo patrón, miro el que tenga mas características iguales

Si tenemos 2 parámetros con la misma cantidad de características iguales, puedo hacer la media de las y para sacar el resultado

Si no encaja ningún patrón, calculo la distancia y tendré que seleccionar la distancia menor

Este algoritmo no tiene un modelo (formula)

Aprendizaje vago, cada vez que tenga una nueva consulta, tendrá que mirar toda la base de datos

Diseño del algoritmo:

- 1.-tengo que categorizar los items
- 2.-calcular la distancia entre los dos patrones (la distancia euclidia) y las sumo
- 3.- calculo la raíz cuadrada

con esto soy capaz de sacar la distancia entre un patrón y otro

k-nn = k vecinos próximos

problemas:

1.- Dimensionalidad

si es de alta dimensión, puede no llegar a tener vecinos

Esto se llama la maldición de la dimensionalidad

2.- Calibración

Un regla eurística popular es $K = \text{raiz de } n$

3.- Escalas de medida (No mirar la tabla)

No es lo mismo usar m o ft

Cuando uno corre el algoritmo en otro país puede dar un resultado erróneo por la escala de medida

Para resolverlo, escalamos todo y listo

4.- La independencia de las características

puede que una características tenga correlación con otra pero con el modelo se está contando 2 veces

ejm: peso y altura (las personas altas tendrán más peso que las bajas)

5.- importancia de las características

no todas las características tienen el mismo peso (Nivel de renta)

para solucionarlo: puedo ponderar las características

6.- Sensibilidad frente a características influyentes

7.-Complejidad computacional y altísimo coste de memoria