



Ricardo Aler Mur

AUTOEVALUACIÓN DE CLASIFICADORES KNN-1, CON PREGUNTAS Y RESPUESTAS

- 1) ¿Por qué los clasificadores basados en vecino(s) más cercano(s) funcionan mal en presencia de atributos irrelevantes?

Respuesta: porque se basan en la idea de cercanía, la cual se mide con una distancia. Por ejemplo, si usamos la distancia Euclídea, esta distancia se calcula como la suma de varias diferencias al cuadrado, tantas como atributos. Si muchos de los atributos son irrelevantes, muchos de los componentes de la distancia serán arbitrarios y también lo será la clasificación que resulte de ella.

- 2) ¿En qué se convierte KNN si K =número de datos de entrenamiento?

Respuesta: se convierte en un clasificador trivial que siempre clasifica con la clase mayoritaria de los datos. Más en detalle, sin K =número de datos, todos los datos forman parte de la vecindad del dato a clasificar, y eso es así independientemente del dato que queramos clasificar. Por tanto este clasificador, siempre responderá con la misma clase, la clase mayoritaria presente en los datos de entrenamiento.

- 3) ¿Cómo podemos disminuir la influencia de los datos con ruido en la clasificación de KNN?

Respuesta: de varias maneras. Podemos utilizar $K > 1$. También podemos eliminar las instancias ruidosas con algún algoritmo de selección de instancias, en concreto algoritmos de edición, como la regla de edición de Wilson

- 4) ¿Qué ventaja tiene la regla de Reduced Nearest Neighbor frente a la de Condensed Nearest Neighbor?

Respuesta: Además de eliminar datos redundantes como CNN, también es capaz de quitar datos ruidosos.

- 5) Decir una ventaja de LVQ frente a KNN

Respuesta: LVQ es mucho más rápido a la hora de clasificar nuevos datos, puesto que sólo tiene que comprobar la distancia del dato a unos pocos prototipos, mientras que KNN tiene que comprobar la distancia a todos los datos de entrenamiento.