

implementacion del algoritmode clasificacion k-nn

Luis Francisco Matlalcuatzi González



6 de febrero de 2024

Benemerita universidad autonoma de puebla

Facultad en Ciencias de la Computación

El algoritmo de los vecinos más cercanos, también conocido como k-NN (k-nearst neighbors), es un algoritmo utilizado en aprendizaje automático y reconocimiento de patrones. Se utiliza para clasificar objetos en función de los rasgos o características que poseen.

En el algoritmo k-NN, se toma un conjunto de datos de entrenamiento etiquetados, es decir, cada dato tiene una clase conocida. Luego, para clasificar un nuevo dato, se busca en el conjunto de entrenamiento los k vecinos más cercanos al nuevo dato, basándose en alguna métrica de distancia, como la distancia euclidiana. Una vez encontrados los k vecinos más cercanos, se asigna la clase que más se repite entre estos vecinos al nuevo dato.

La elección de k es un parámetro importante en el algoritmo de k-NN. Un valor pequeño de k puede llevar a una clasificación ruidosa, ya que se dará más peso a los pocos vecinos más cercanos. Por otro lado, un valor grande de k puede generar una clasificación demasiado generalizada, ya que se promediarán muchas clases diferentes.

El algoritmo de k-NN es considerado un algoritmo de "aprendizaje perezoso" porque posterga el cálculo del modelo hasta el momento de la clasificación. No se producen cambios en el modelo durante la fase de entrenamiento, sino que todo el trabajo se realiza en la etapa de clasificación.

En resumen, el algoritmo de k-NN se basa en encontrar los vecinos más cercanos a un nuevo dato para clasificarlo en función de la clase más común entre estos vecinos. Es un algoritmo sencillo pero efectivo, especialmente en conjuntos de datos pequeños. Sin embargo, su rendimiento puede verse afectado por la elección de k y la elección de la métrica de distancia adecuada.