

Algoritmo de aprendizaje de la máquina KNN

Daniela Monserrat García Sotelo

A01365499

Jorge Abraham Sánchez Mora

A01364653

K-Nearest - Neighbor (Vecinos más cercanos)

Es un algoritmo basado en instancias de tipo supervisado de Machine Learning. También puede usarse para clasificar nuevas muestras o valores discretos para predecir la regresión, valores continuos. Sirve mucho para clasificar los valores más similares y también se basa en conjeturas de nuevos puntos de esa nueva clasificación.



Funcionamiento del algoritmo

$$D_{m,n} = \begin{bmatrix} d_1 = & a_{1,1} & \dots & a_{1,n} & c_1 \\ d_2 = & a_{2,1} & \ddots & a_{2,n} & c_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ d_m = & a_{m,1} & \dots & a_{m,n} & c_m \end{bmatrix}$$

1. Calcula la distancia entre el item a clasificar y el resto de los items del dataset.
2. Seleccionar los “k” elementos más cercanos (con menor distancia, según la función que se use)
3. Realizar una “votación de mayoría” entre los k puntos: los de una clase/etiqueta que dominen decidirán su clasificación final

Nota: El valor k es muy importante ya que gracias k se define a qué grupo pertenecen los puntos sobre todo en las fronteras entre grupos. Por ejemplo, se pueden elegir los valores impares de k para desempatar (si las features que utilizamos son pares). No será lo mismo tomar 3 datos que 23 (No por tomar más puntos significa que vaya a mejorar la precisión, sino que se tardará más nuestro algoritmo en procesar y darnos respuesta).

Ventajas

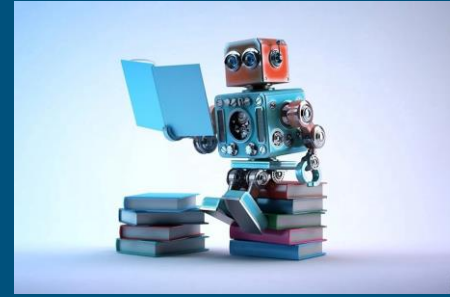
No paramétrico: No hace suposiciones explícitas sobre la forma funcional de los datos, evitando los peligros de la distribución subyacente de los datos.

Algoritmo simple: Es sencillo de aprender, explicar e interpretar.

Alta precisión (relativa): Es bastante alta pero no competitiva en comparación con modelos de aprendizaje mejor supervisados.



Desventajas



Basados en instancia: El algoritmo no aprende explícitamente un modelo, en su lugar, elige memorizar las instancias de capacitación que se utilizan posteriormente como conocimiento para la fase de predicción. Es decir, cuando hacen una consulta a la base de datos para que prediga una etiqueta dada una entrada, el algoritmo usará las instancias de entrenamiento para escupir una respuesta.

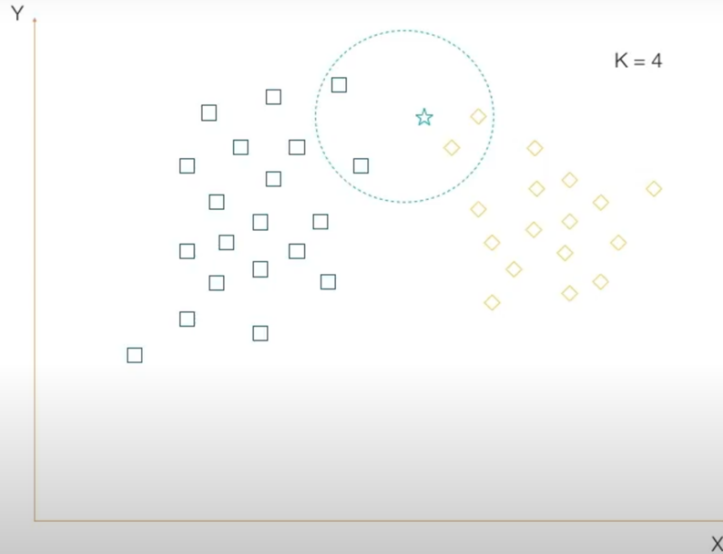
Computacionalmente costoso: Porque el algoritmo almacena todos los datos de entrenamiento.

Requisito de memoria alta: Almacena todos (o casi todos) los datos de entrenamiento.

Problemas en los que se aplica (regresión, clasificación)

APRENDIZAJE SUPERVISADO: K-NEAREST NEIGHBORS | #11 Curso de Introducción a Machine Learning

Aprendizaje Supervisado: K-Nearest Neighbors

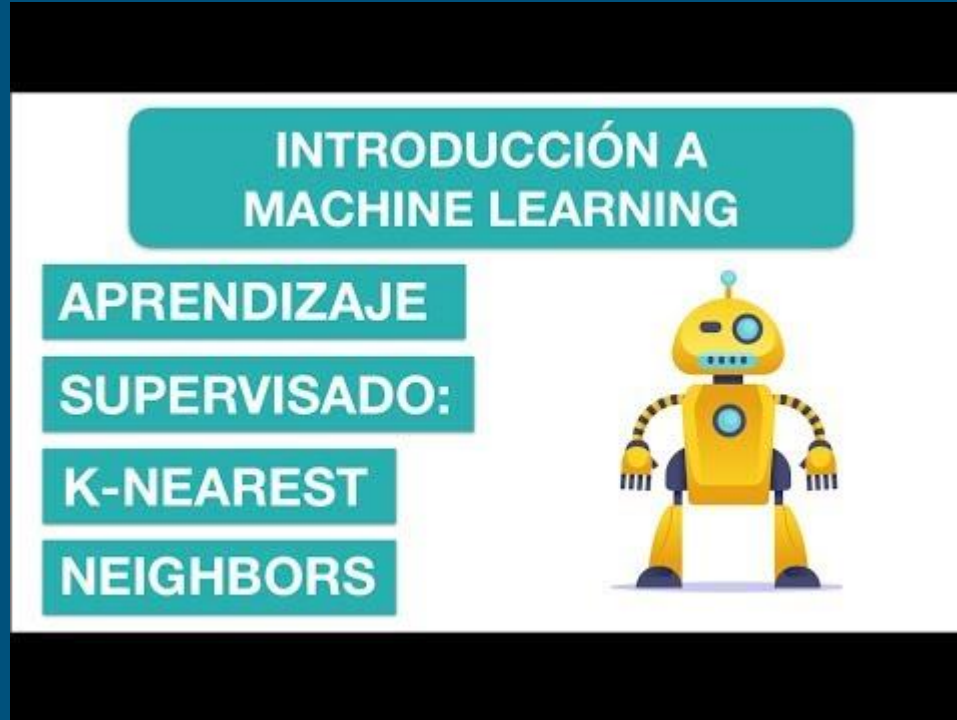


▶ ⏮ 🔊 3:33 / 6:06

Desliza hacia abajo para ver más detalles



Ejemplo



Bibliografía:

S.A.. (2018). Aprende Machine Learning. 26/09/2020, de S.A. Sitio web: <https://www.aprendemachinelearning.com/clasificar-con-k-nearest-neighbor-ejemplo-en-python/#:~:text=K%2DNearest%2DNeighbor%20es%20un,el%20mundo%20del%20Aprendizaje%20Autom%C3%A1tico>.

Lidgi Gonzalez. (2018). Aprende todo sobre inteligencia artificial. 27/09/2020, de S.A. Sitio web: <https://ligdigonzalez.com/aprendizaje-supervisado-k-nearest-neighbors/>