

Introducción a K Vecinos más cercanos

K Nearest Neighbors (KNN)

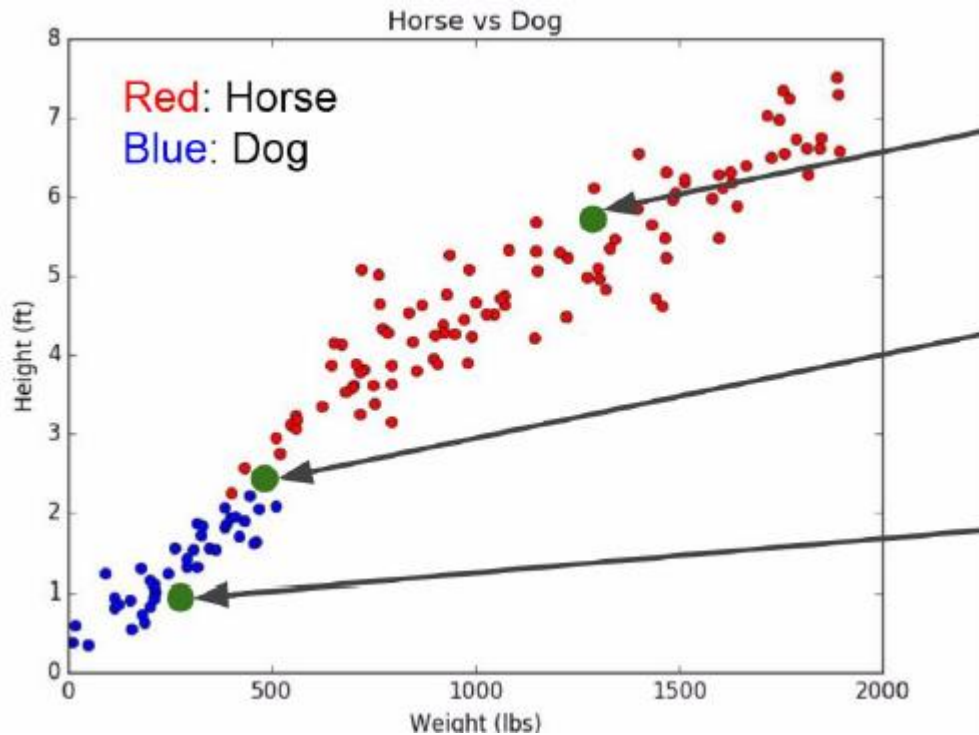
Lectura Sugerida

Capítulo 4 de
Introduction to Statistical Learning
Gareth James

KNN

- K Nearest Neighbors es un algoritmo de clasificación que opera sobre un principio muy simple.
- ¡Se muestra mejor a través de un ejemplo!
Imagina que tenemos algunos datos imaginarios sobre perros y caballos, con alturas y pesos.

KNN



Nuevo punto de datos:
¿Es un caballo o un perro?

Nuevo punto de datos:
¿Es un caballo o un perro?

Nuevo punto de datos:
¿Es un caballo o un perro?

KNN

Algoritmo de entrenamiento:

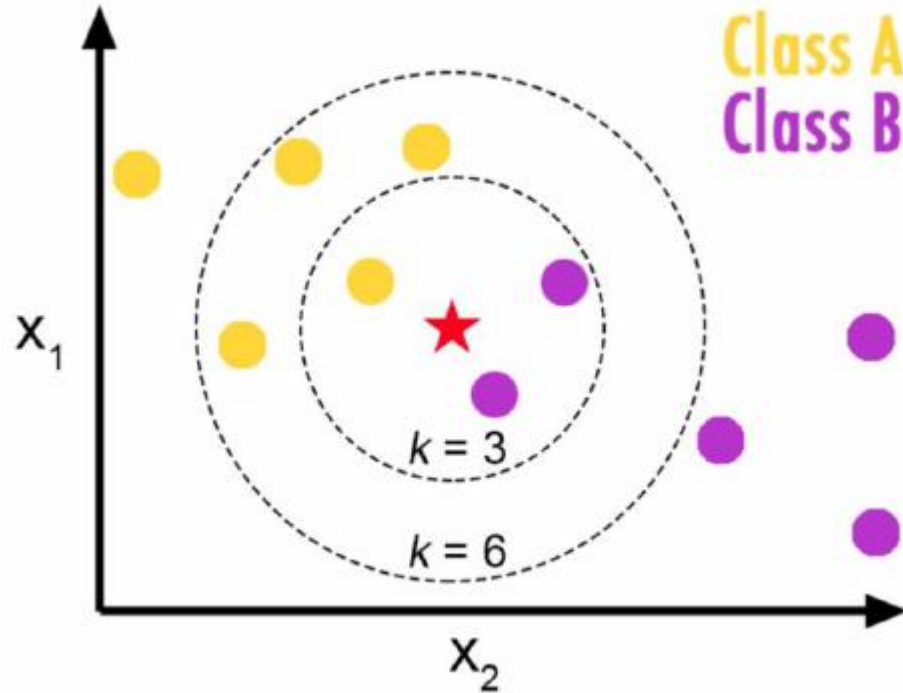
1. Almacenar todos los datos.

Algoritmo de predicción:

1. Calcula la distancia de x a todos los puntos en tus datos
2. Ordena los puntos en tus datos aumentando la distancia de x
3. Predice la etiqueta de la mayoría de los puntos más cercanos a la " k "

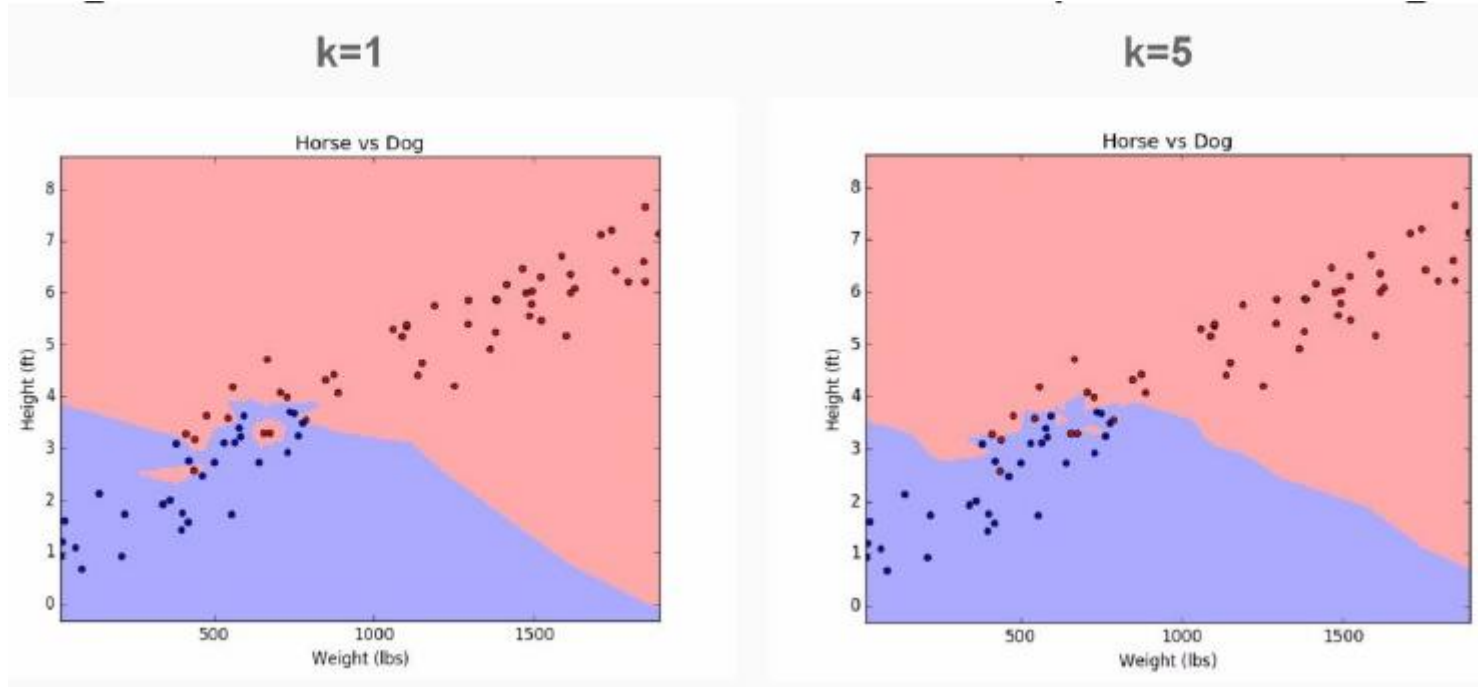
KNN

Elegir una K afectará a la clase a la que se asigna un nuevo punto:



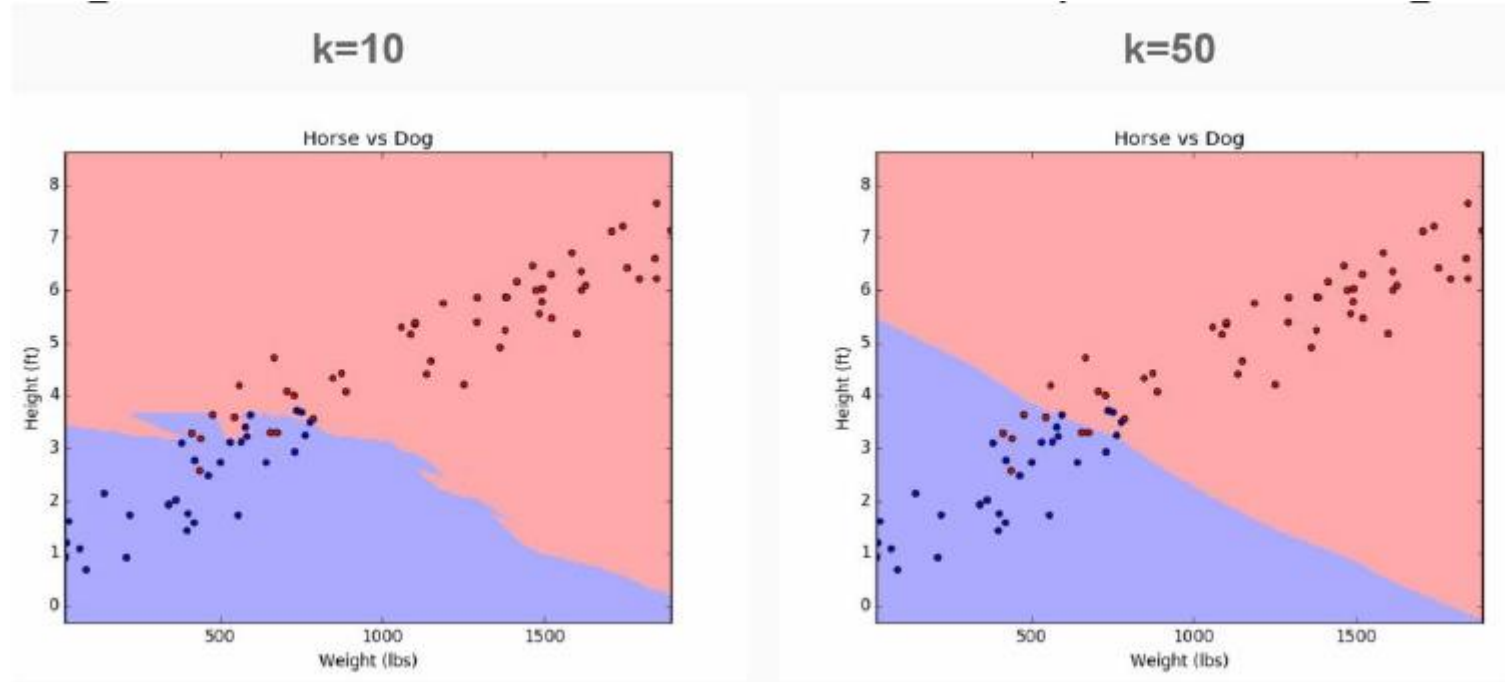
KNN

Elegir una K afectará a la clase a la que se asigna un nuevo punto:



KNN

Elegir una K afectará a la clase a la que se asigna un nuevo punto:



KNN - Pros

- Muy simple
- El entrenamiento es trivial.
- Trabaja con cualquier número de clases.
- Fácil de agregar más datos
- Pocos parámetros
 - ✓ K
 - ✓ Métrica de distancia

KNN Contrás

- Alto costo de predicción (peor para conjuntos de datos grandes)
- No es bueno con datos de altamente dimensionales.
- Las características categóricas no funcionan bien