

Introdução ao Método dos K Vizinhos Mais Próximos (K Nearest Neighbors – KNN)

Atividade de Leitura

Todo o Capítulo 4 do livro
Introduction to Statistical Learning
por Gareth James, et al.

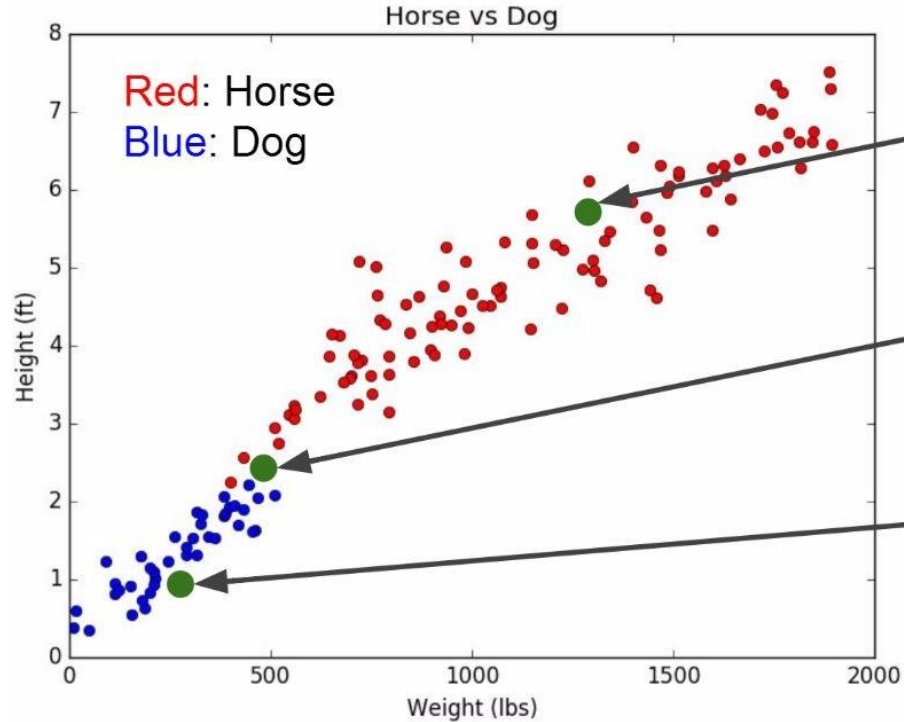
KNN

K Vizinhos Mais Próximos (K Nearest Neighbors) é um algoritmo de **classificação** que opera em um princípio muito simples.

Esse princípio é melhor demonstrado através de um exemplo:

Imagine que tivéssemos alguns dados imaginários sobre Cães e Cavalos, com alturas e pesos.

KNN



New datapoint:
Is it a horse or a dog?

New datapoint:
Is it a horse or a dog?

New datapoint:
Is it a horse or a dog?

KNN

Algoritmo de treinamento:

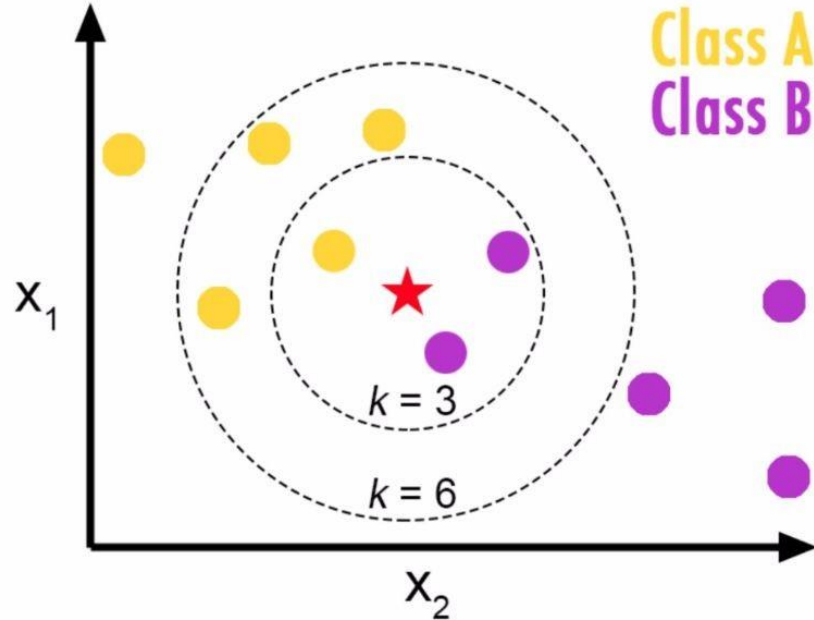
1. Armazene todos os dados

Algoritmo de previsão:

1. Calcule a distância de x a todos os pontos em seu conjunto de dados
2. Organize os pontos em seu conjunto de dados pelo aumento da distância de x
3. Verifique o rótulo majoritário dos “ k ” pontos mais próximos

KNN

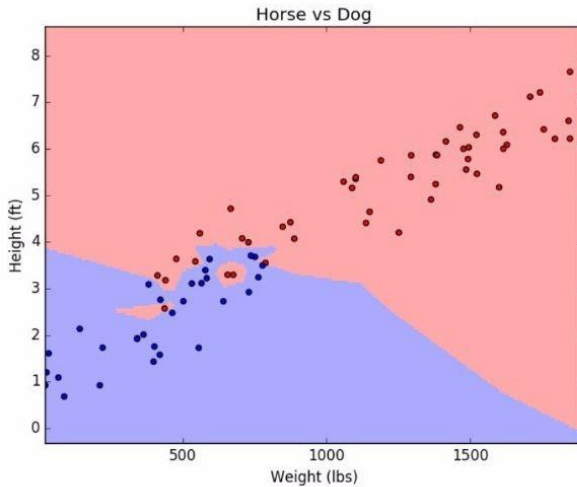
Escolher um valor de K afetará a qual classe um novo ponto é atribuído:



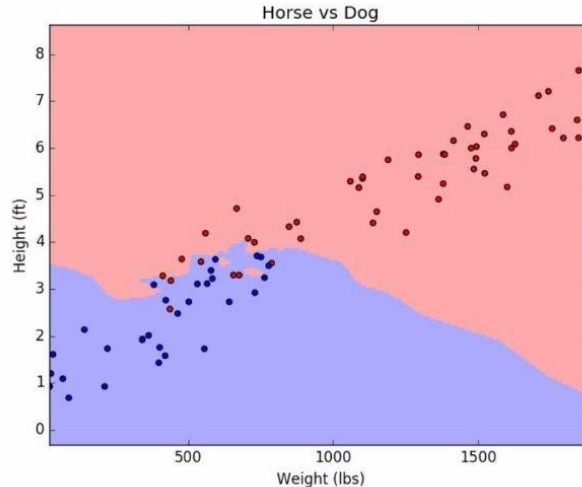
KNN

Escolher um valor de K afetará a qual classe um novo ponto é atribuído:

k=1

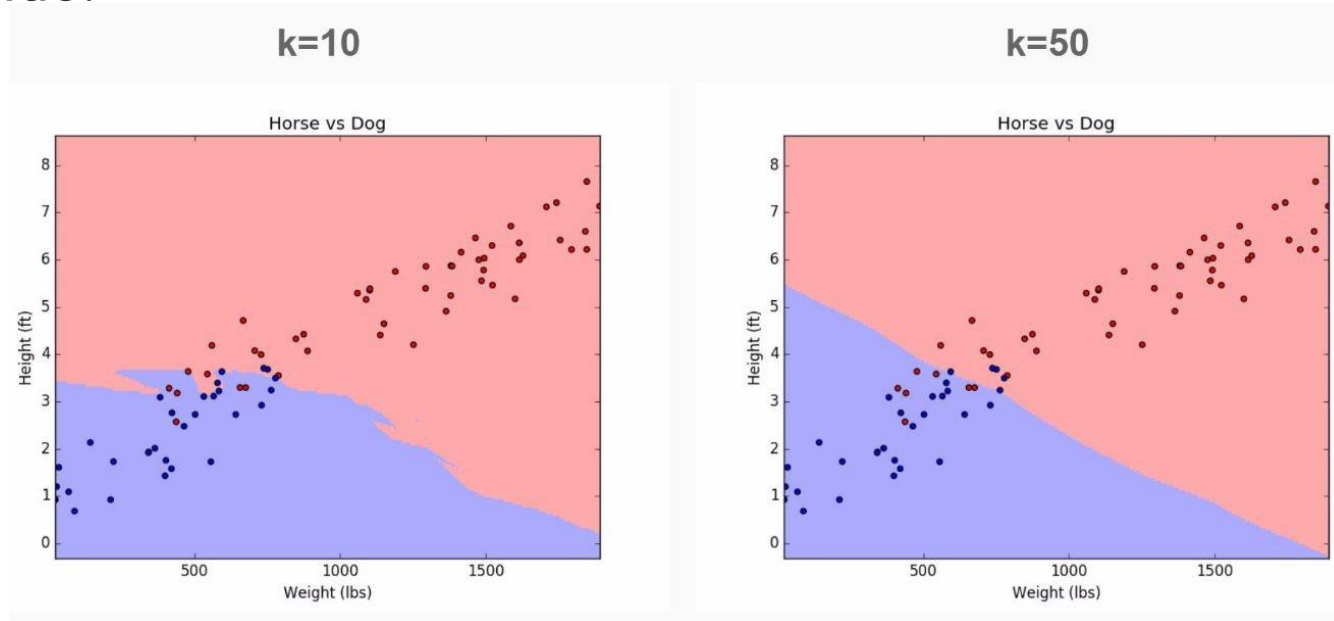


k=5



KNN

Escolher um valor de K afetará a qual classe um novo ponto é atribuído:



KNN

Prós

- Muito simples
- O treinamento é trivial
- Funciona com qualquer número de classes
- É fácil adicionar mais dados
- Poucos parâmetros
 - K
 - Métrica de distância (distance metric)

KNN

Contras

- Alto custo de previsão (pior para grandes conjuntos de dados)
- Não é bom com dados de muitas dimensões
- Características categóricas não funcionam bem