# Introdução a Aprendizagem De Máquina

Pós-graduação em Ciência de Dados e Machine Learning
Módulo 3 - Data Mining e Machine Learning

Professor Msc. Ricardo José Menezes Maia

# K Nearest Neighbors-KNN

# K Nearest Neighbors  KNN

Seção 4.6 do Introduction to Statical Learning de Gareth James

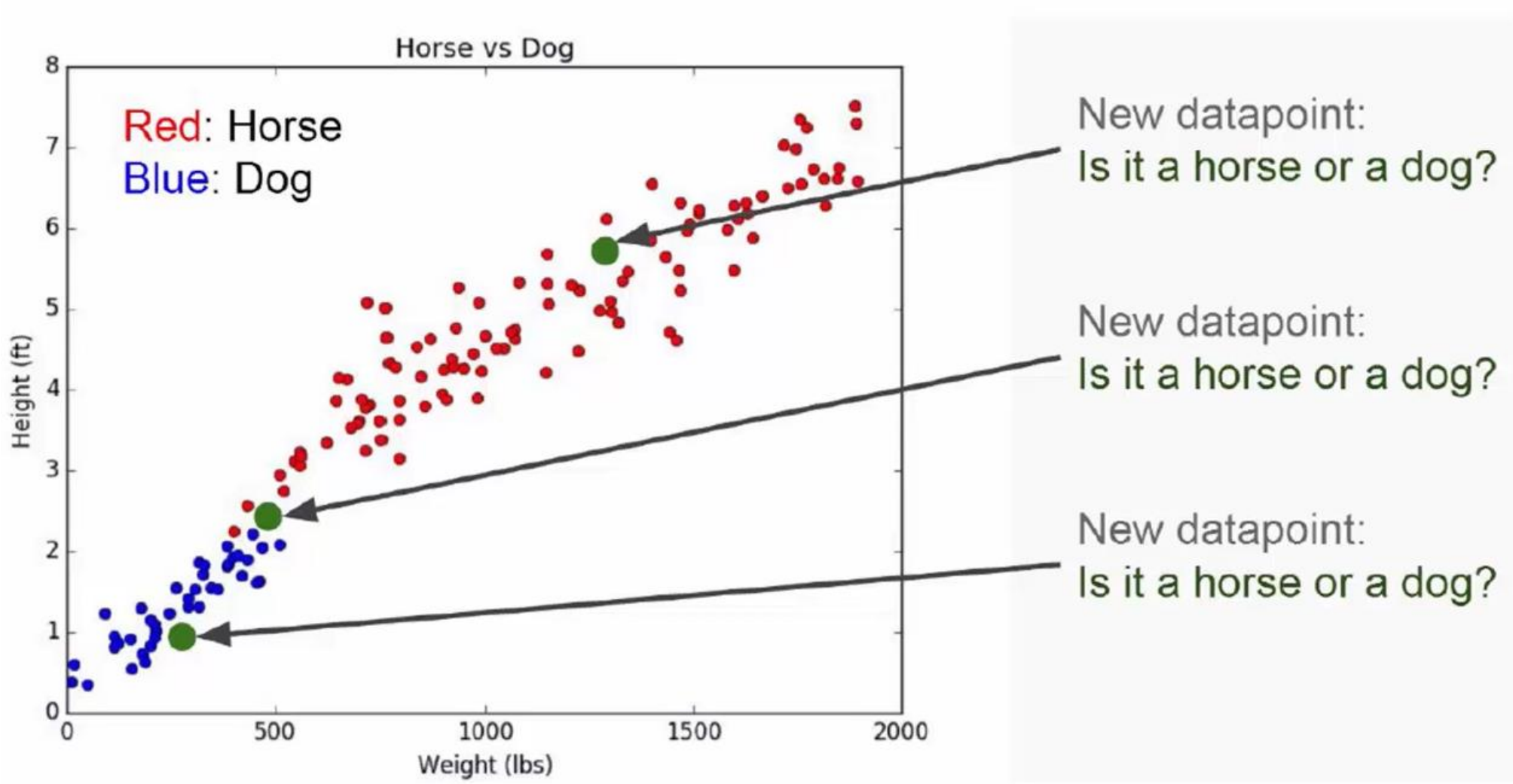
Método dos K vizinhos mais próximos

KNN é um algoritmo de classificação de dados que opera de forma muito simples.

Vamos explicar  seu funcionamento com um exemplo!

Imagine que temos alguns dados imaginários de alturas e pesos de cachorros e cavalos

# K Nearest Neighbors  KNN



# K Nearest Neighbors KNN

Algoritmo de treino:

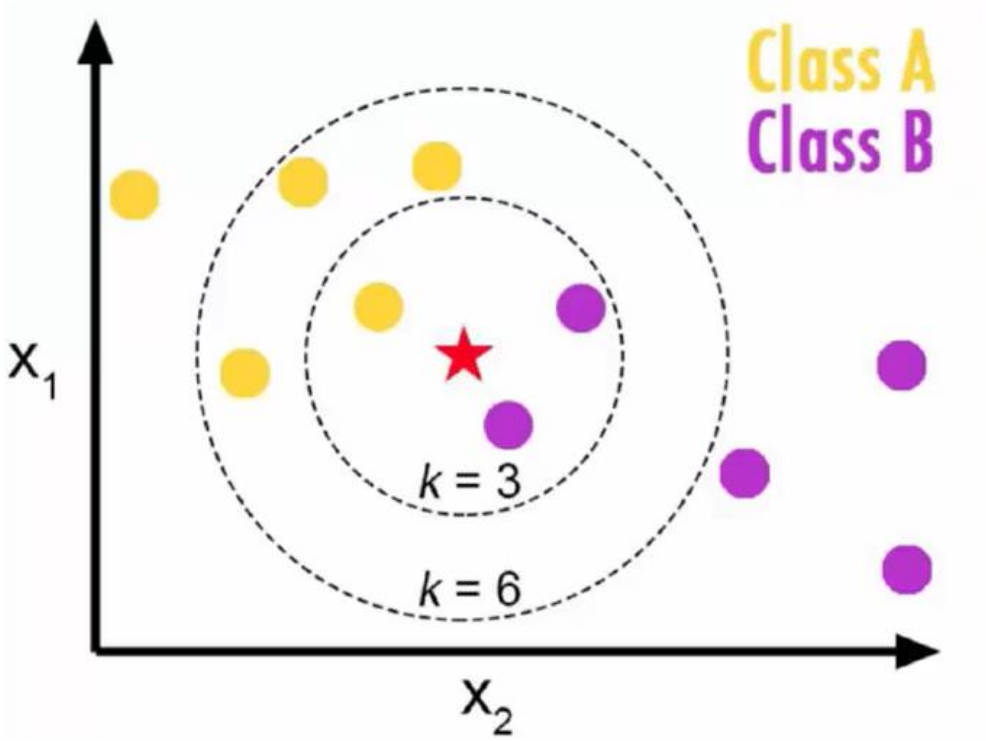
1. Guarde os dados

Algoritmo de teste/preditor:

1. Calcule as distâncias do x até os demais pontos.
2. Organize os dados em ordem crescente de distância.
3. Classifique a classe de acordo com a maioria dos primeiros ‘K’ valores

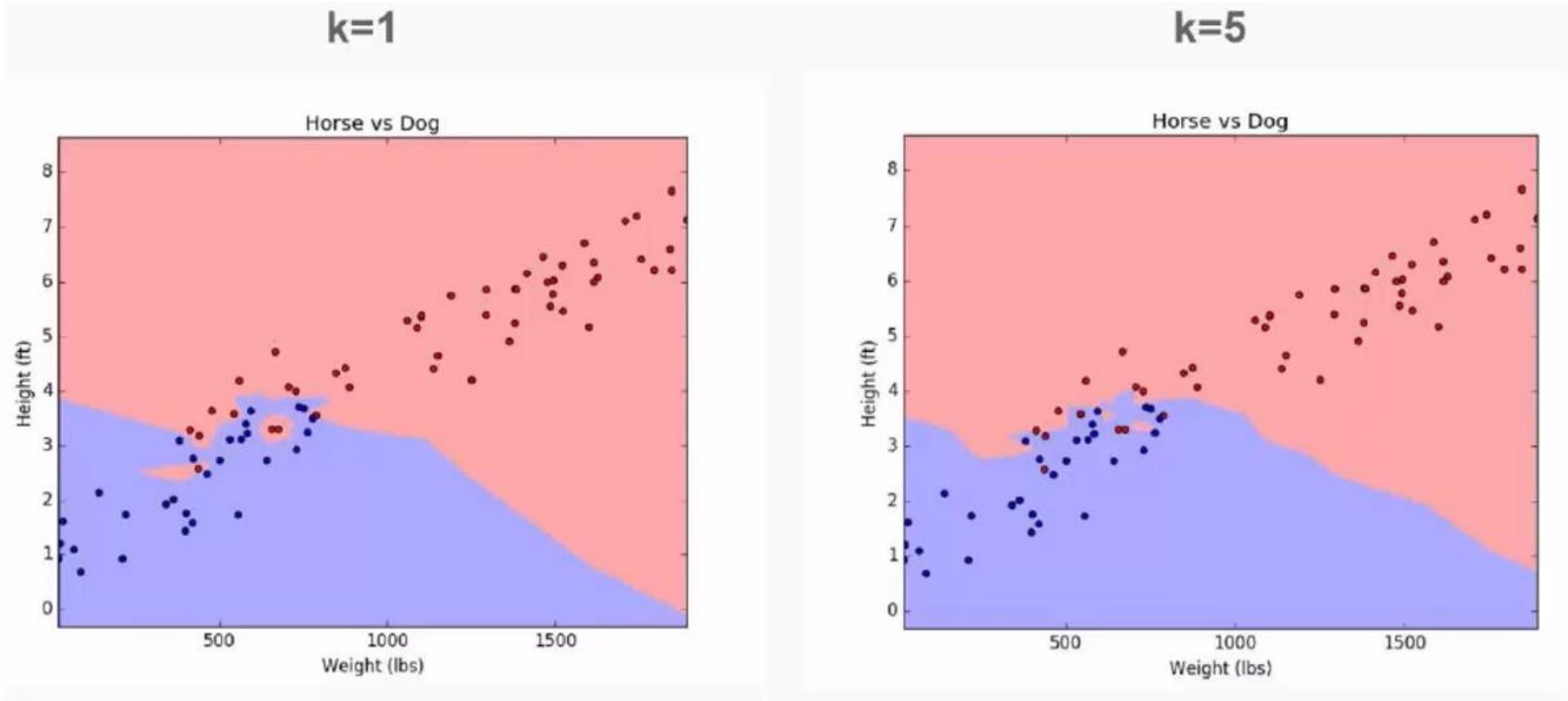
# K Nearest Neighbors KNN

O parâmetro K pode afetar a classificação do mesmo:



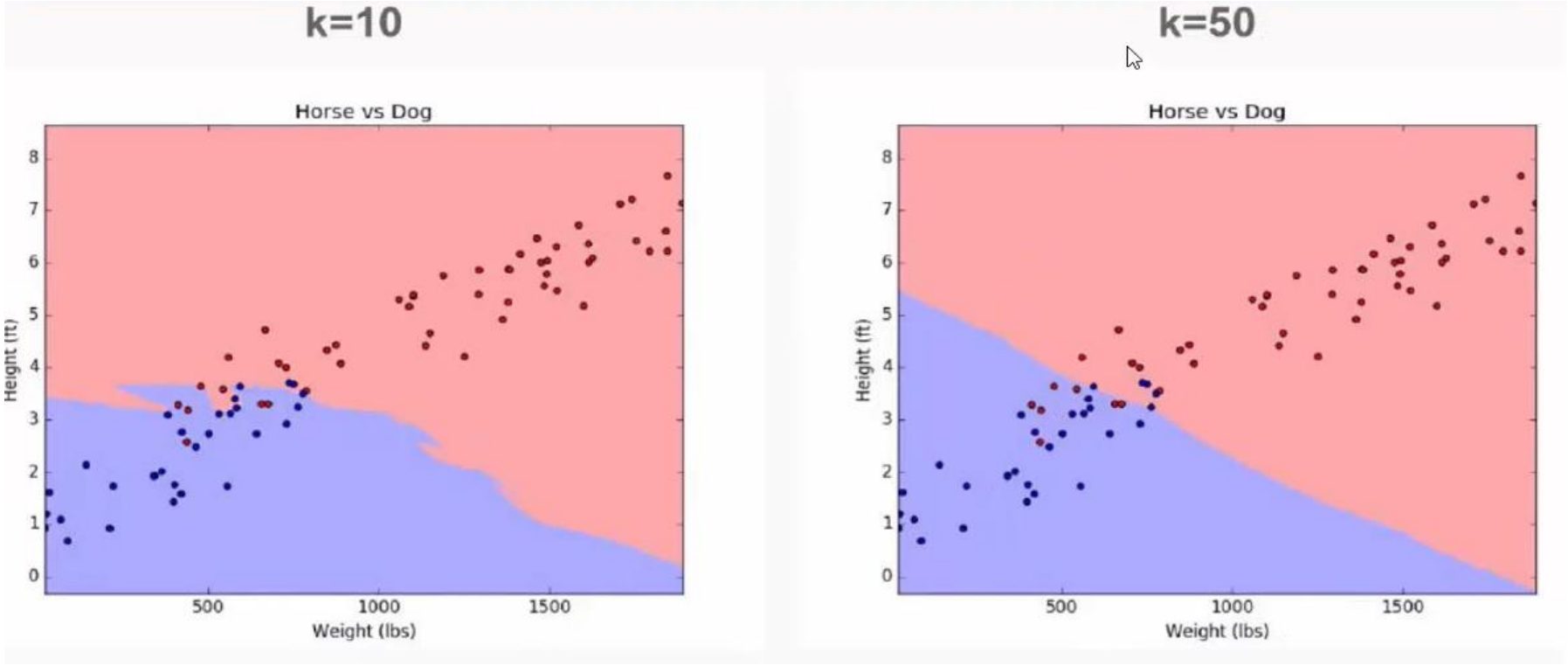
# K Nearest Neighbors KNN

O parâmetro K pode afetar a classificação do mesmo:



# K Nearest Neighbors KNN

O parâmetro K pode afetar a classificação do mesmo:
Para k=50 temos bias muito maior e variância muito menor.
É importante encontrar uma relação onde o K seja estável e faça sentido para nosso modelo.



# K Nearest Neighbors KNN

Pros:

- Muito simples
- Processo de treino é trivial
- Funciona muito bem com um grande número de classes.
- Fácil de se adicionar dados.
- Poucos parâmetros (K e  métrica de distância).

Contras:

- Elevado custo computacional para predição (pior para grandes conjuntos de dados)
- Não muito bom em dados com múltiplas dimensões (muitos parâmetros)
- Parâmetros categóricos não funcionam muito bem.