

# K-ésimos vizinhos mais próximos K Nearest Neighbors (KNN)

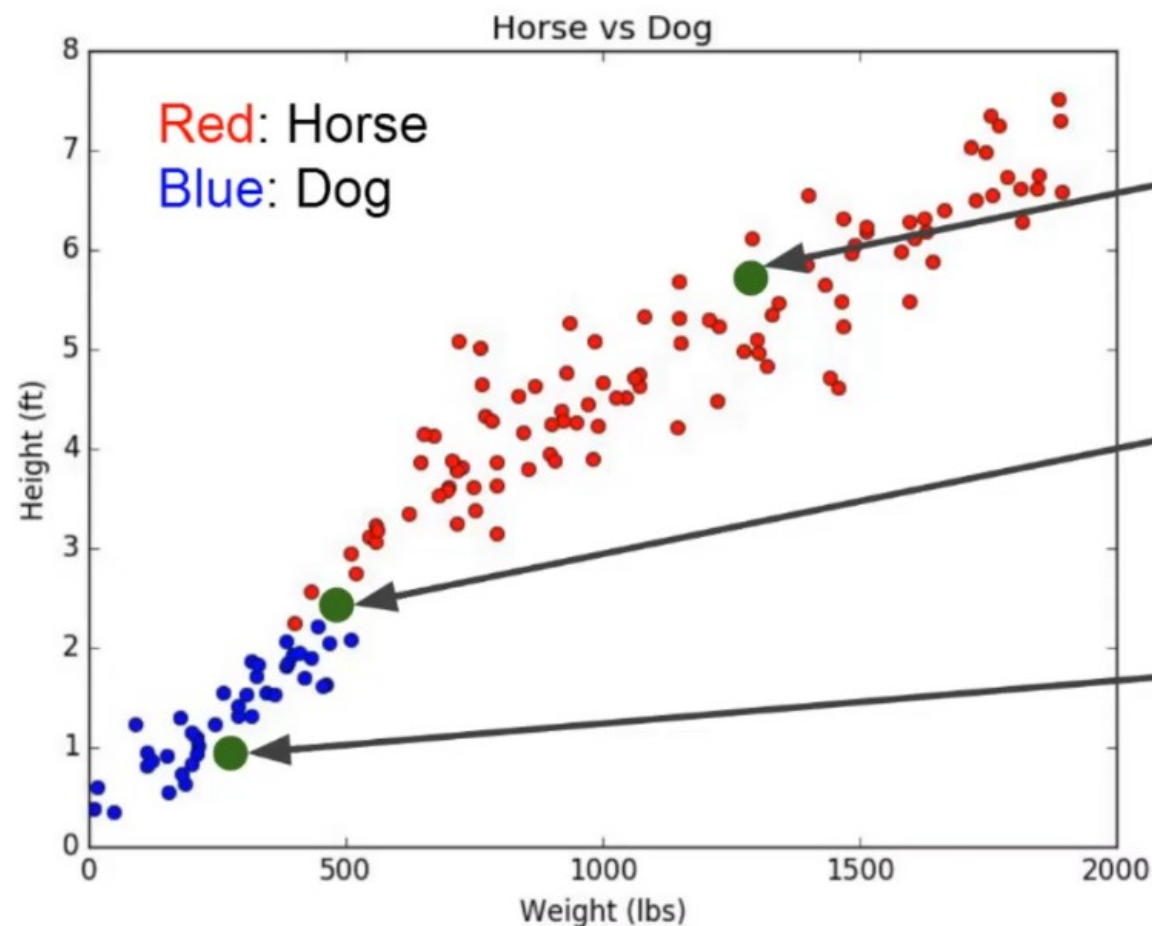
**Instituto Federal de São Paulo - IFSP**

**Campus Guarulhos**

# K-ésimos vizinhos mais próximos

- KNN é um algoritmo de classificação de dados que opera de forma mais intuitiva.
- Vamos explicar seu funcionamento com um exemplo.
- Imagine que temos alguns dados de altura e peso de animais.

# K-ésimos vizinhos mais próximos



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?

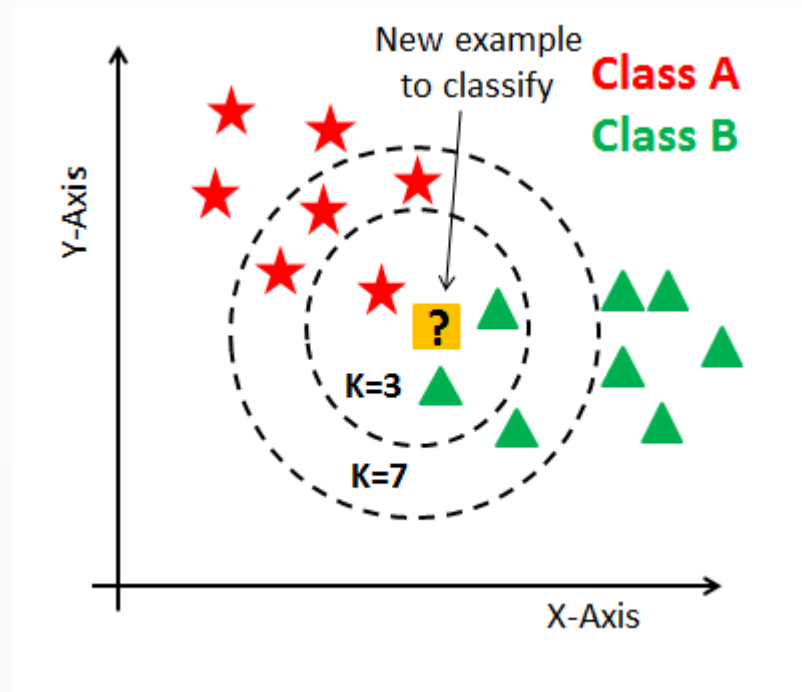
Em cada um dos pontos, quais são cavalos e quais são cachorros ?

# K-ésimos vizinhos mais próximos

- Algoritmo de treino:
  1. Guarde os dados.
- Algoritmo preditor (de teste):
  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' vizinhos mais próximos.

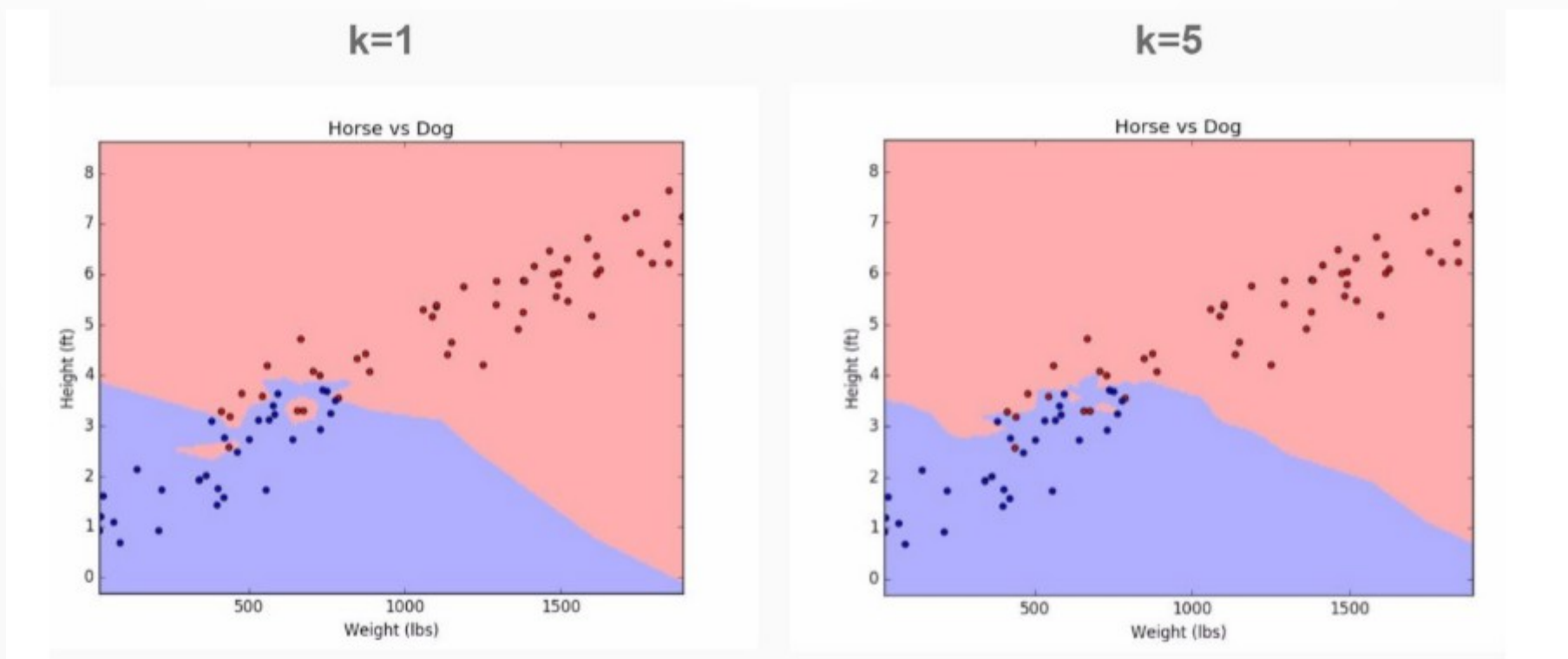
# K-ésimos vizinhos mais próximos

- O parâmetro 'k' pode afetar a classificação do elemento.
- Corresponde ao principal ponto de incerteza no processo de classificação.



# K-ésimos vizinhos mais próximos

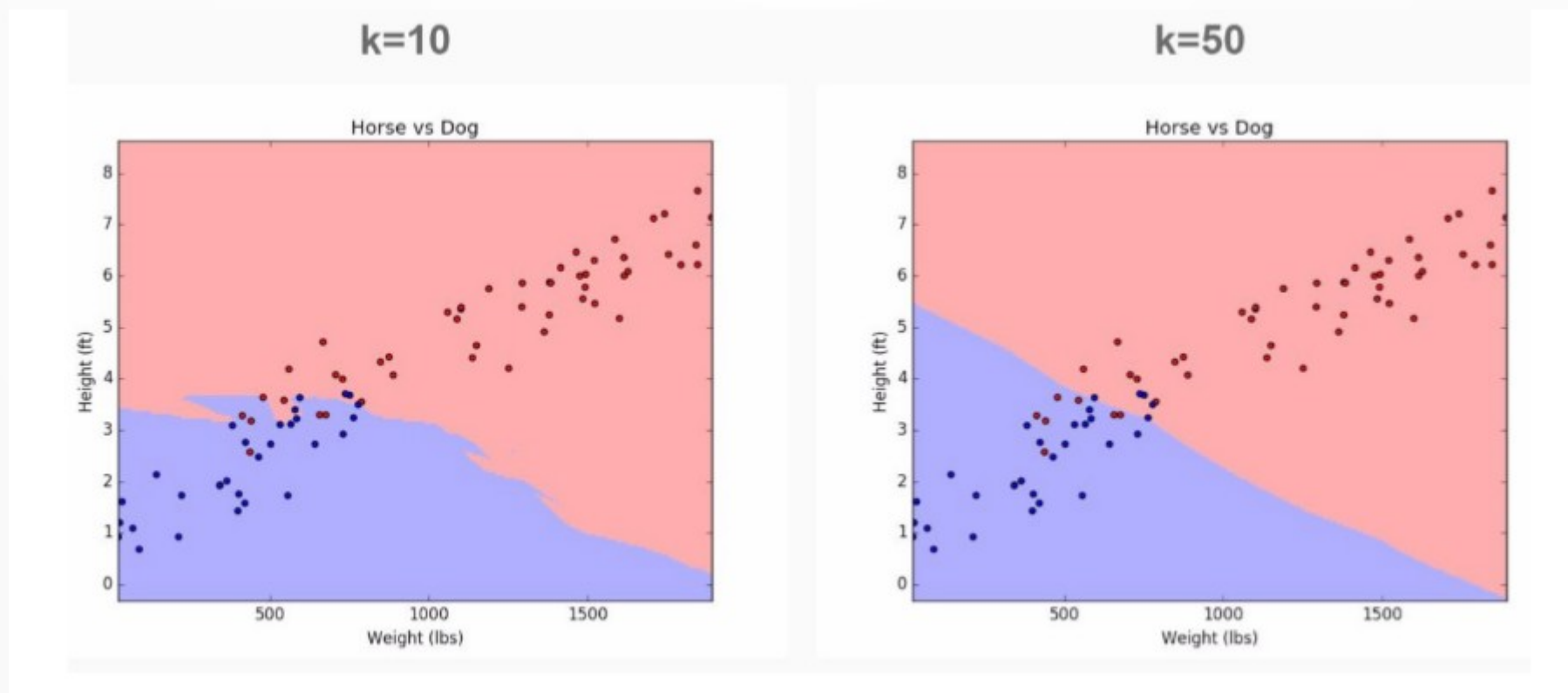
- Como o parâmetro 'k' pode afetar o processo de classificação.





# K-ésimos vizinhos mais próximos

- Como o parâmetro 'k' pode afetar o processo de classificação.



# Pontos positivos do KNN

- Simplicidade;
- Processo de treinamento simples;
- Funciona bem com um grande número de classes;
- Fácil de adicionar mais dados;
- Poucos parâmetros para alteração ( $k$  e métrica de distância).



# Pontos negativos do KNN

- Alto custo computacional para predição (pior para grandes conjuntos de dados);
- Não muito bom em dados com múltiplas dimensões (muitos parâmetros);
- Não funciona bem com parâmetros categóricos.

**Dúvidas!**