# Introduction to K Nearest Neighbors

PIERIAN DATA
www.pieriandata.com

# Reading Assignment

Complete Chapter 4

**Introduction to Statistical Learning**

By Gareth James, et al.

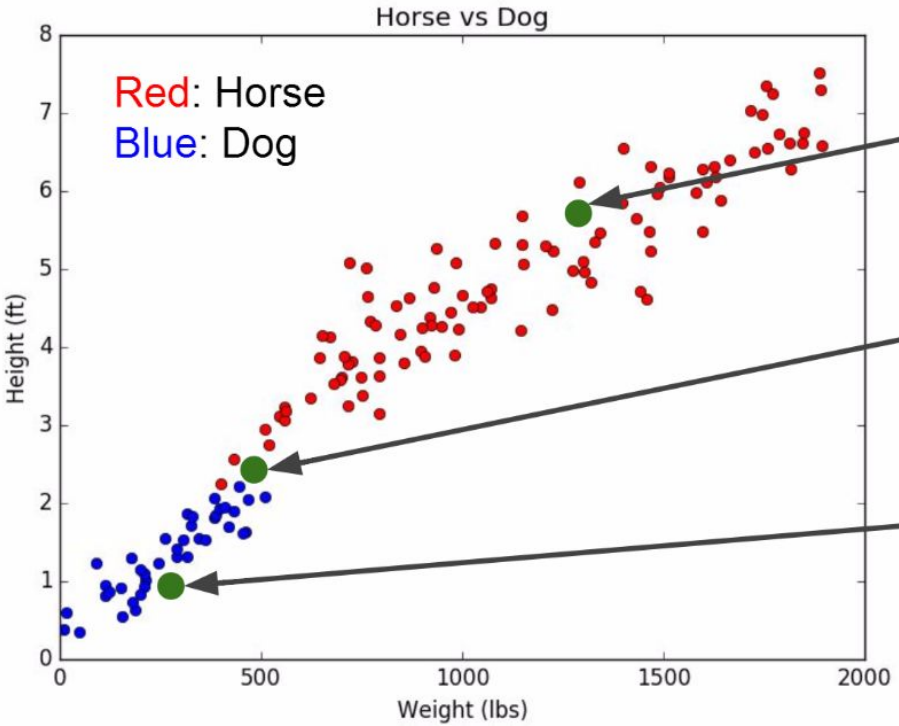
# KNN

K Nearest Neighbors is a **classification** algorithm that operates on a very simple principle.

It is best shown through example!

Imagine we had some imaginary data on Dogs and Horses, with heights and weights.

# 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

Training Algorithm:

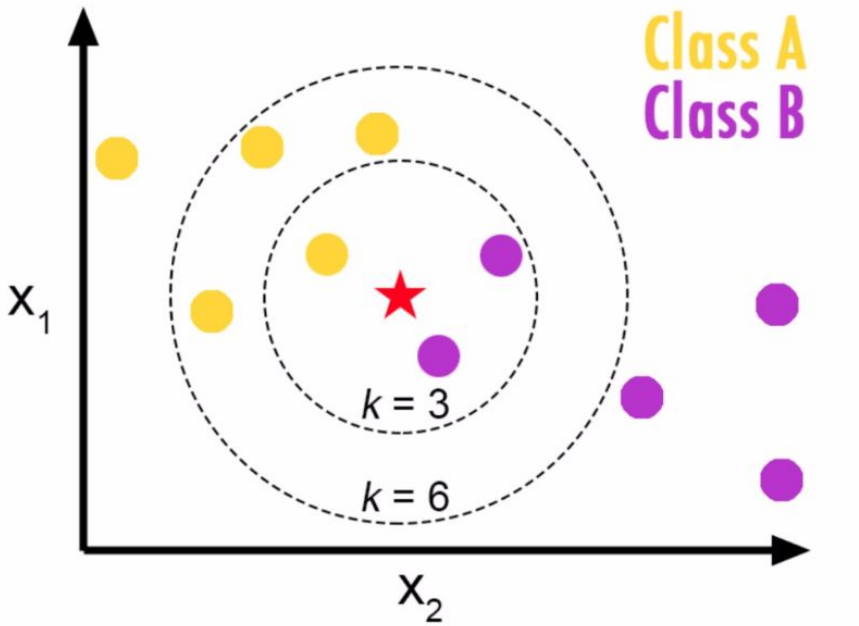
1. Store all the Data

Prediction Algorithm:

1. Calculate the distance from x to all points in your data
2. Sort the points in your data by increasing distance from x
3. Predict the majority label of the “k” closest points

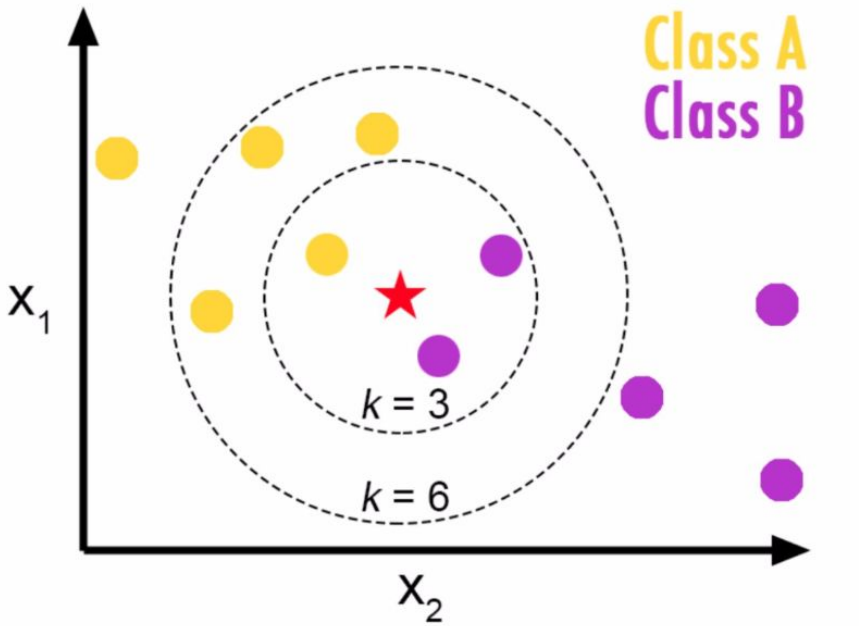
# KNN

Choosing a K will affect what class a new point is assigned to:



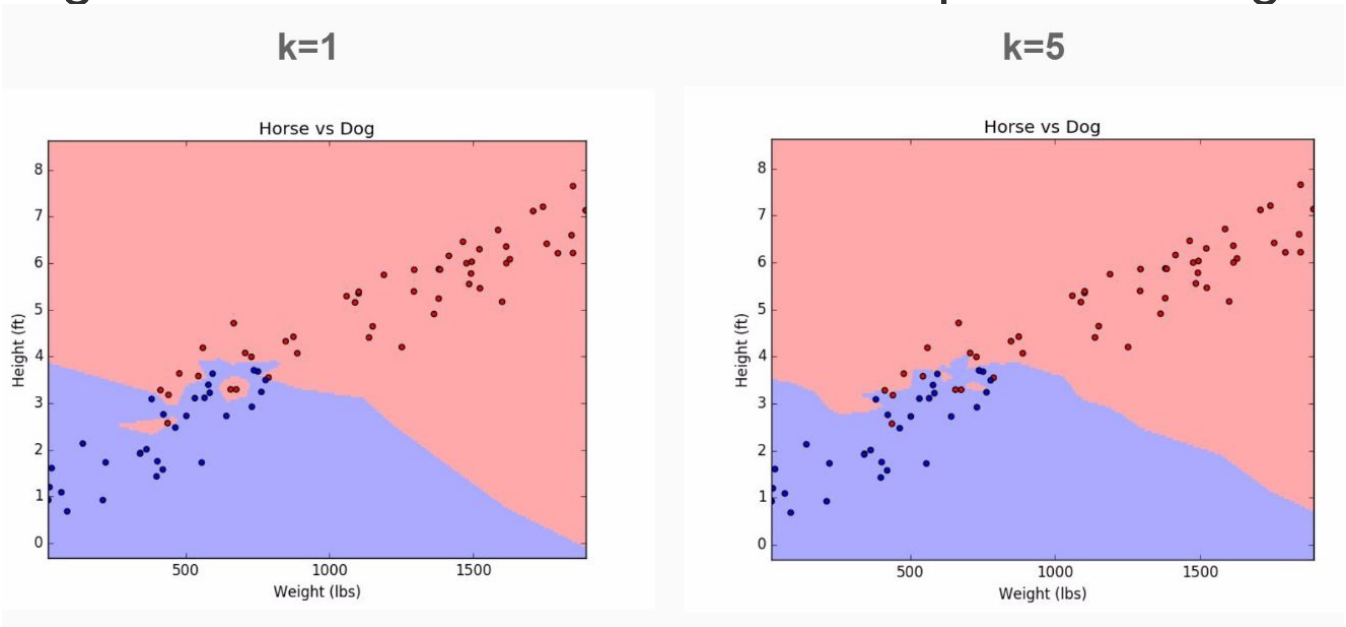
# KNN

Choosing a K will affect what class a new point is assigned to:



# KNN

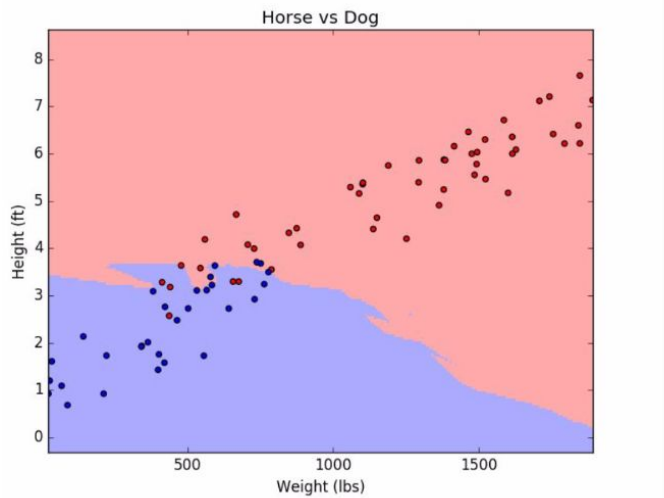
Choosing a K will affect what class a new point is assigned to:



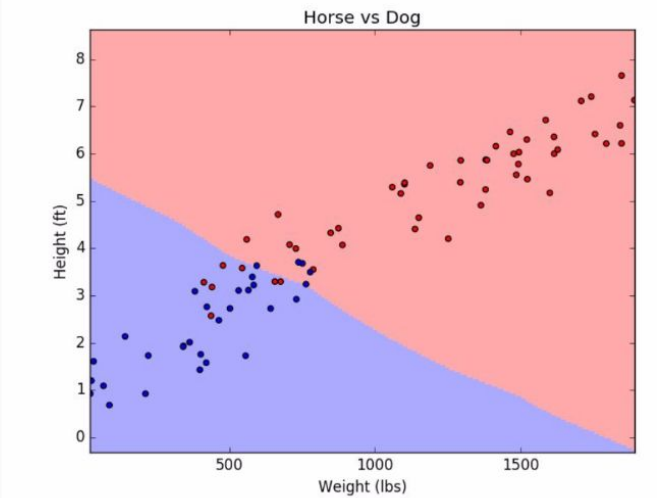
# KNN

Choosing a K will affect what class a new point is assigned to:

k=10




k=50



# KNN

## Pros

- Very simple
- Training is trivial
- Works with any number of classes
- Easy to add more data
- Few parameters
  - K
  - Distance Metric

# KNN

## Cons

- High Prediction Cost (worse for large data sets)
- Not good with high dimensional data
- Categorical Features don’t work well