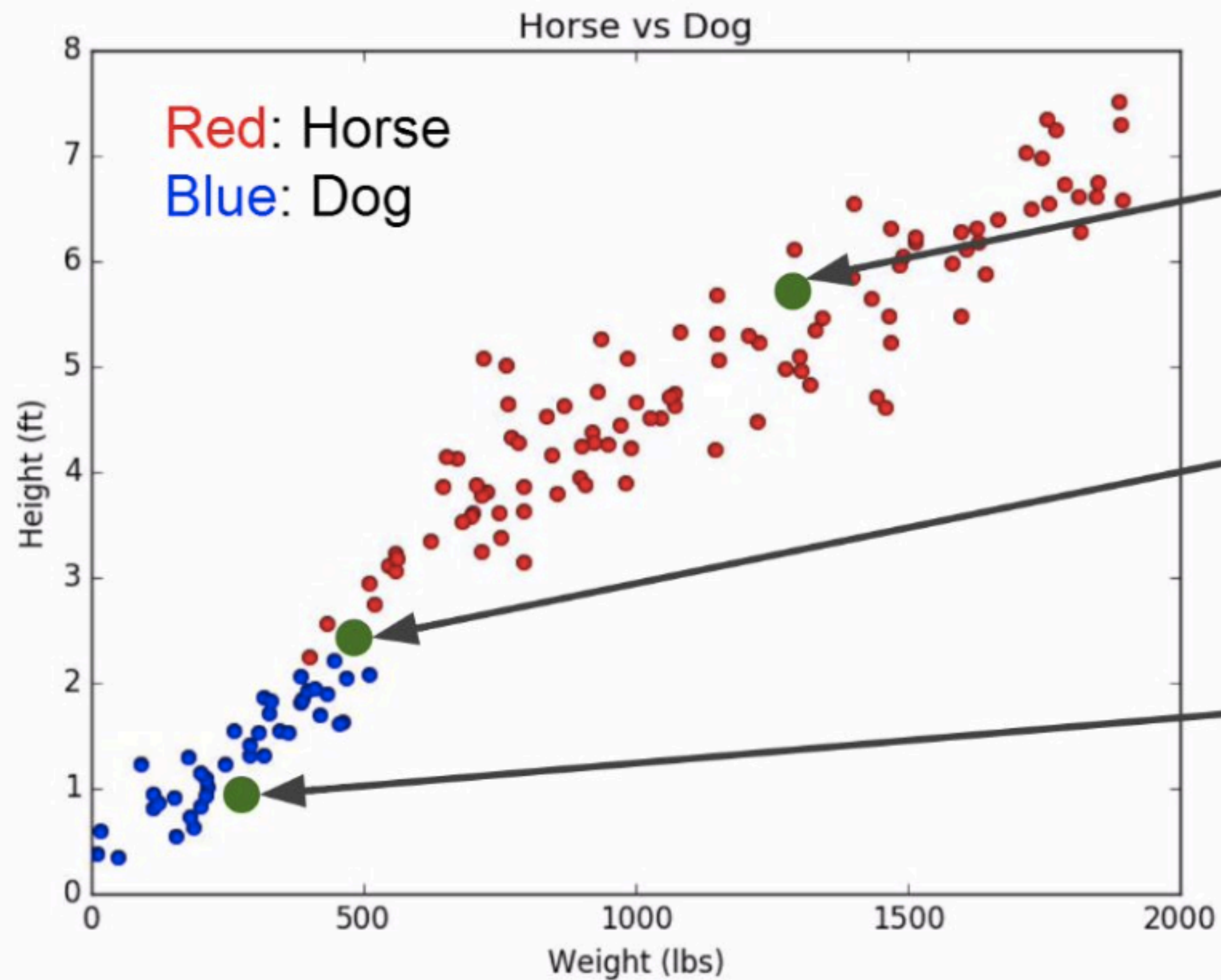


# K-Nearest Neighbours

**K-Nearest Neighbours is a classification algorithm that operates on a very simple principle**



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?

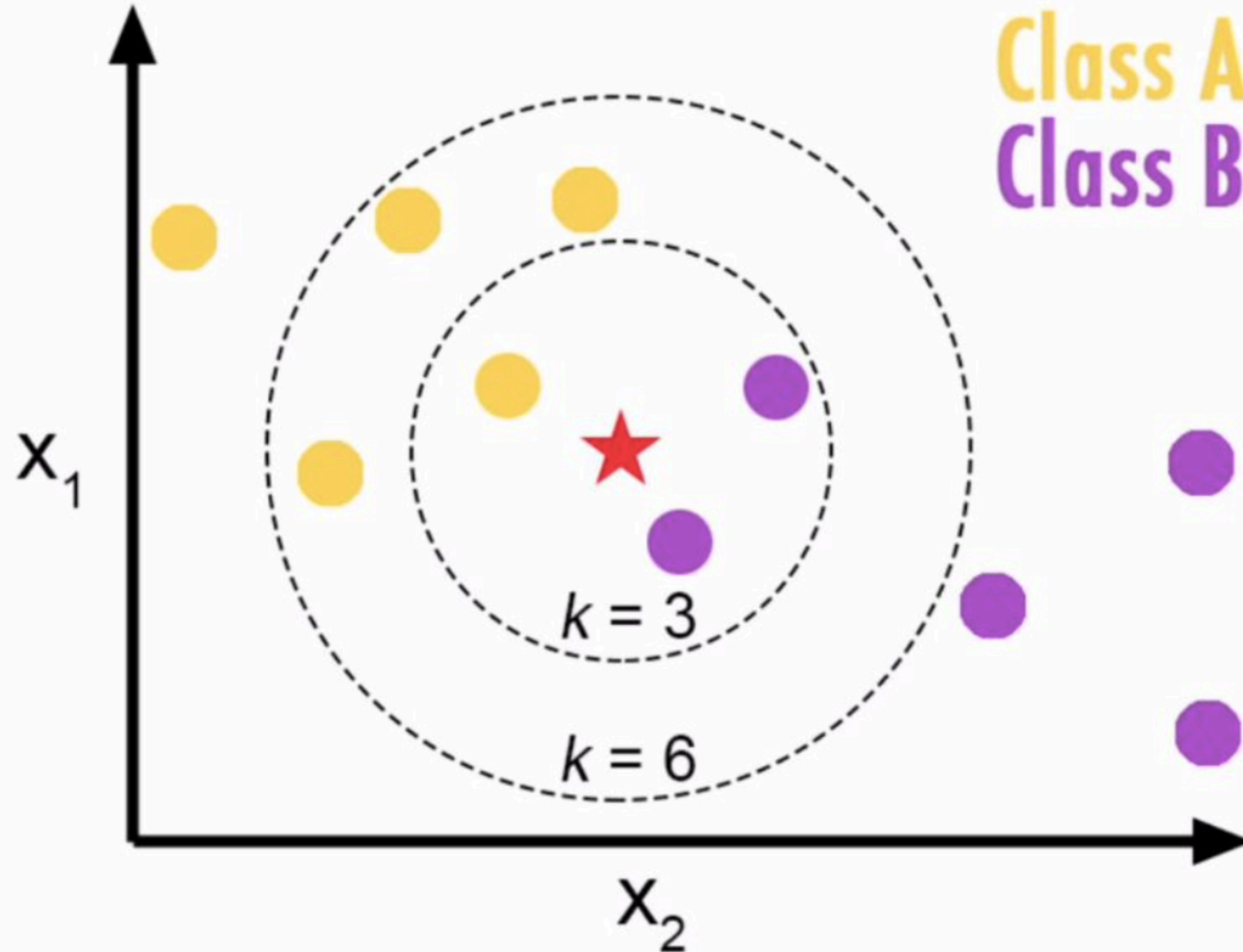
## Training:

- Store all the data

## Prediction:

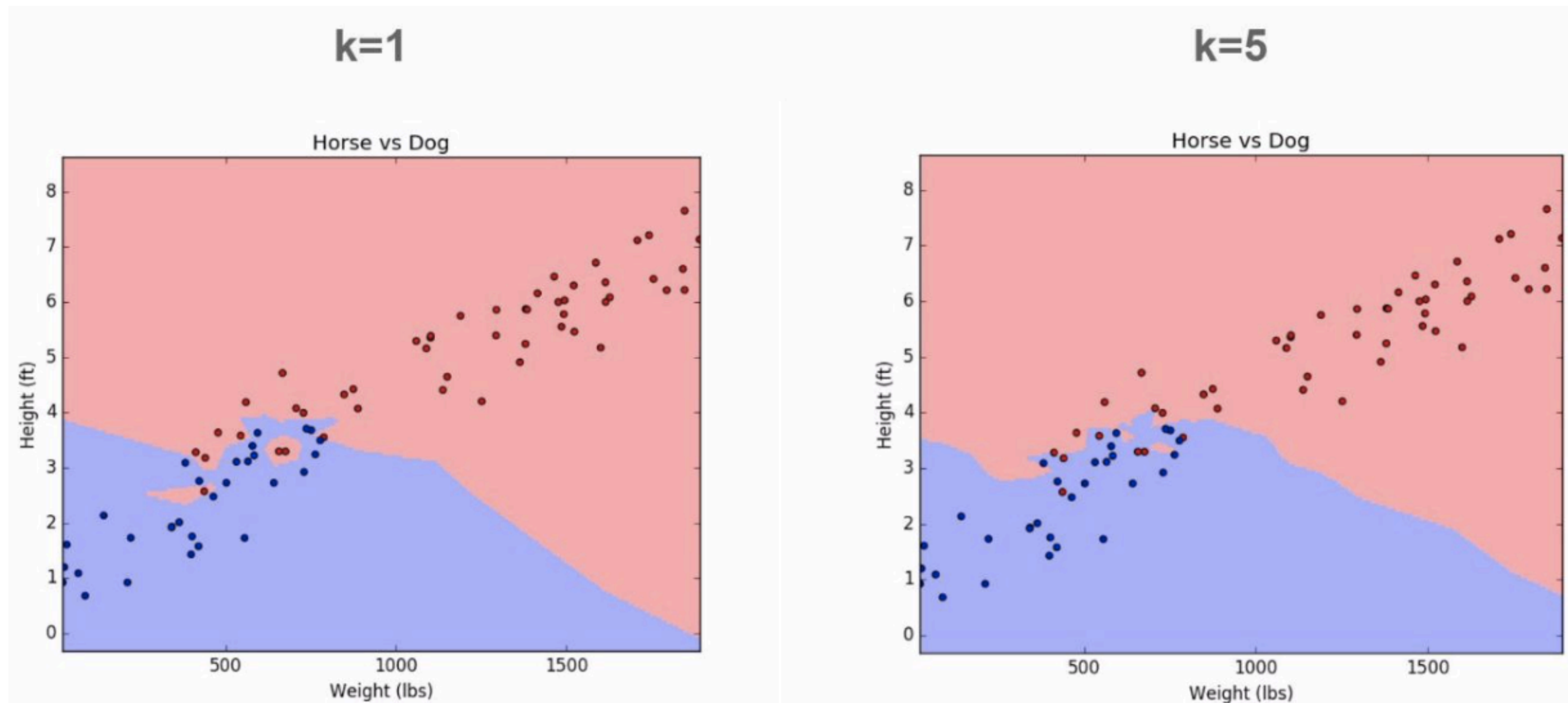
- Calculate the distance from  $x$  to all the points in your data
- Sort the points in your data by increasing distance from  $x$
- Predict the majority label of the  $k$  closest points

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



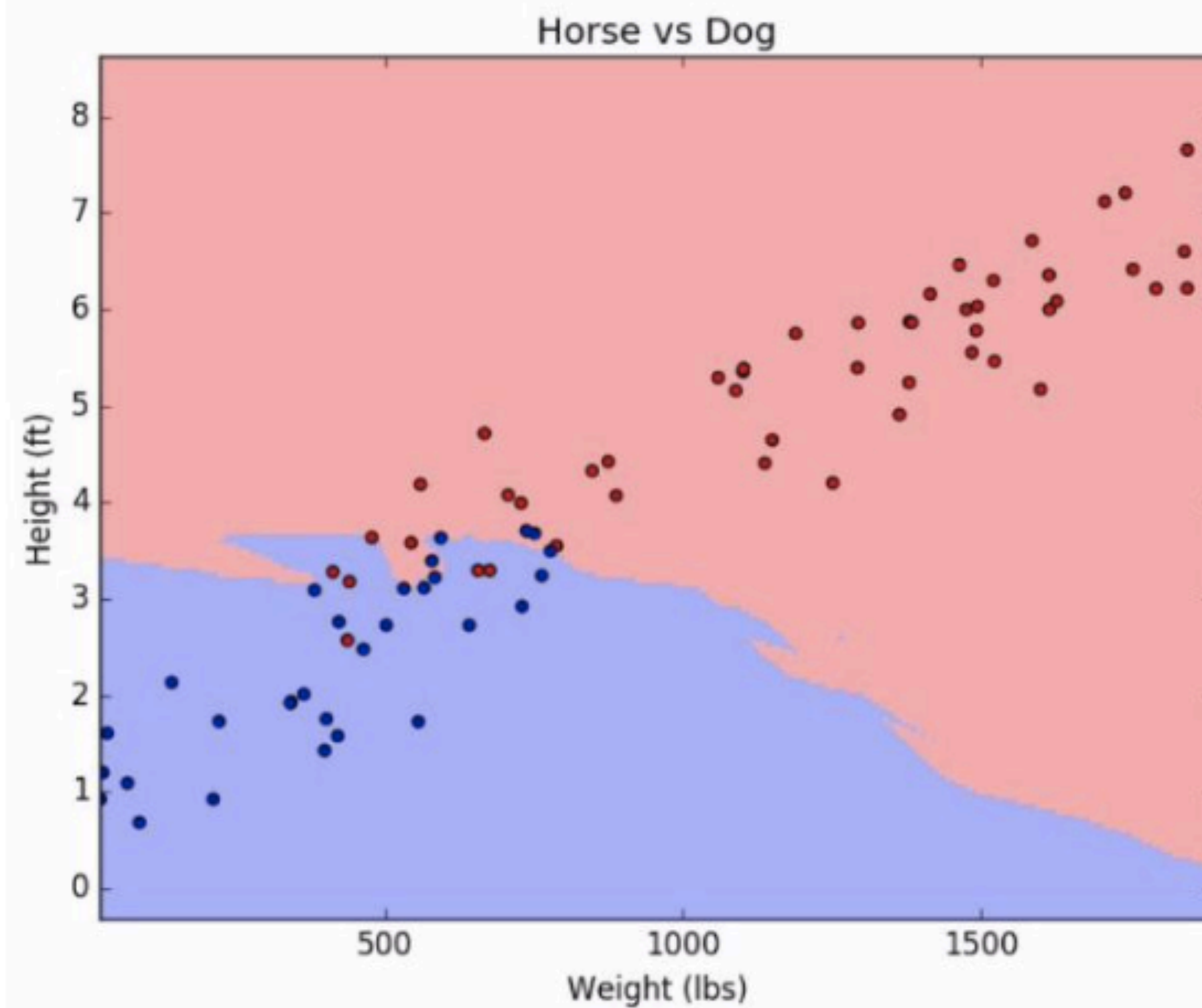


Choosing a K will affect what class a new point is assigned to:

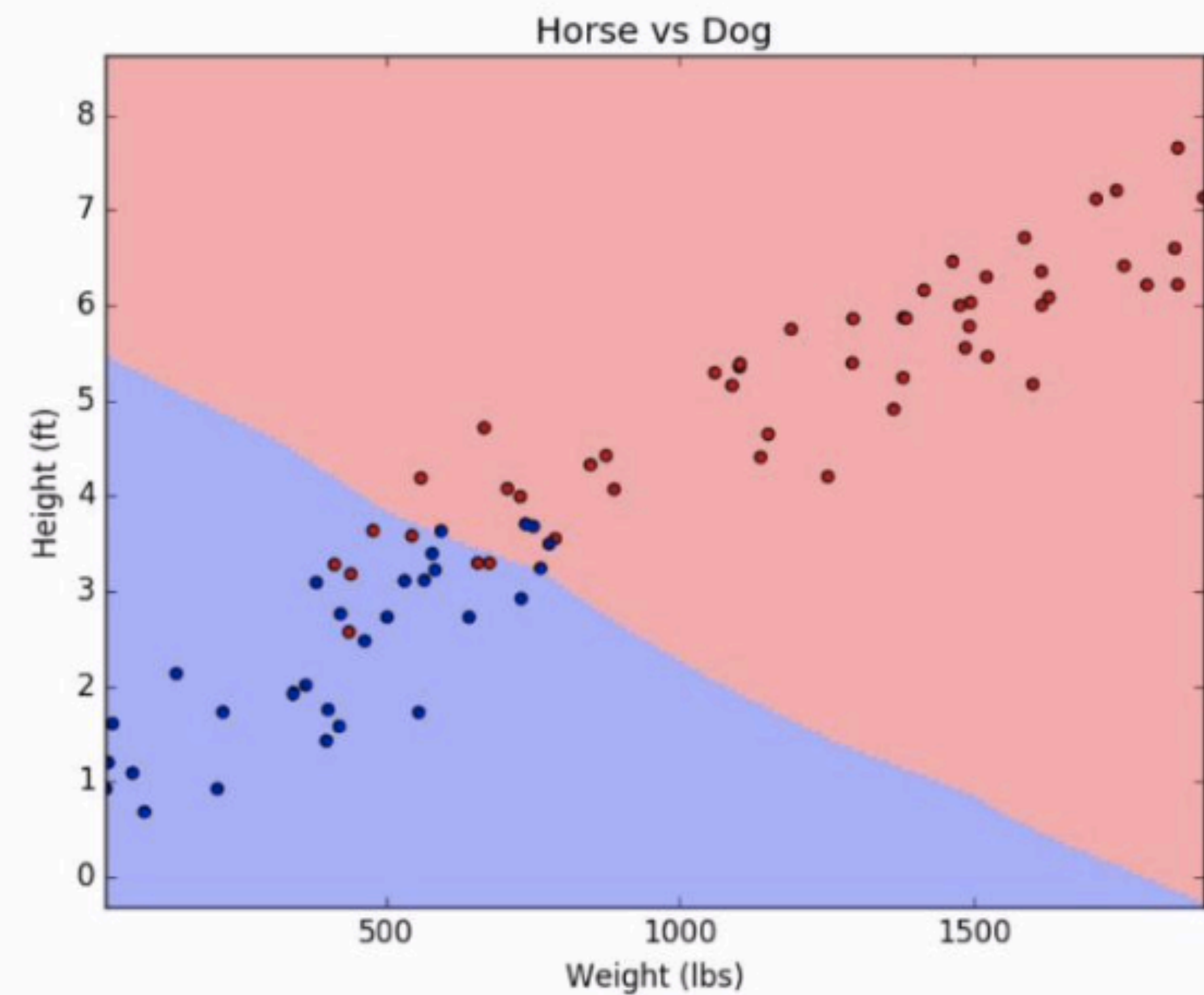


Choosing a K will affect what class a new point is assigned to:

**k=10**



**k=50**



# Pros

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



# Cons

- High prediction cost
- Not good with high dimensional data
- Categorical features don't work well