

Introduction to KNN

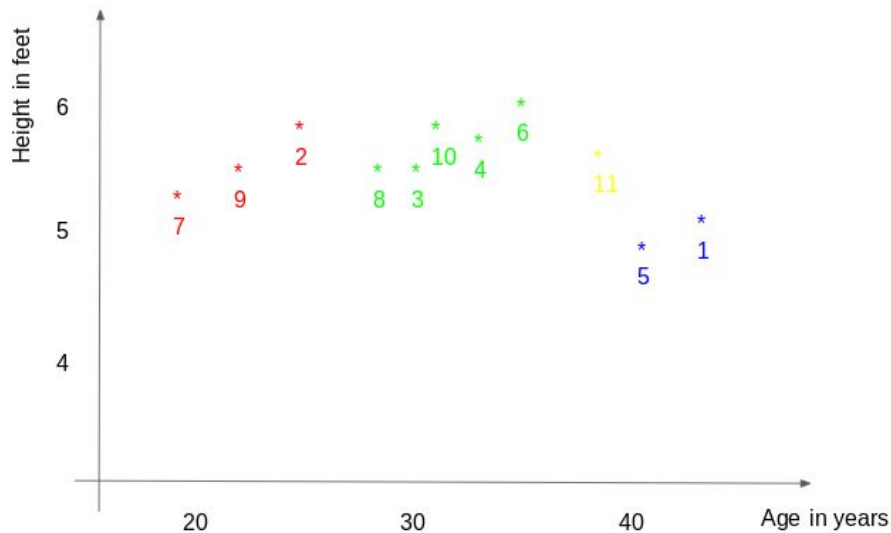


Warm Up

Can you think of any other ways (other than linear regression) to predict a new value based on previously collected data?

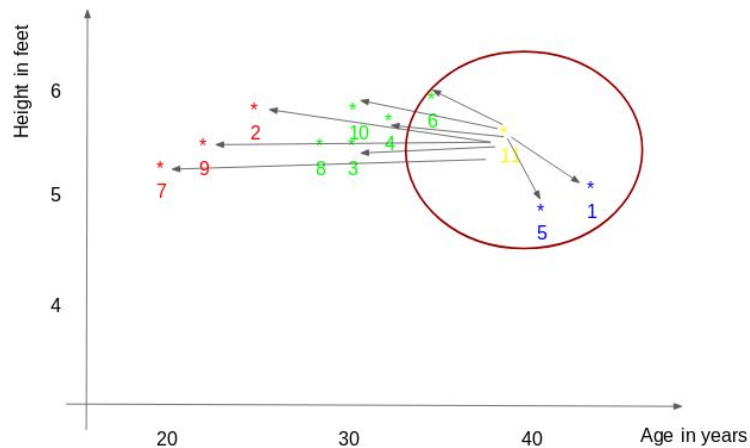
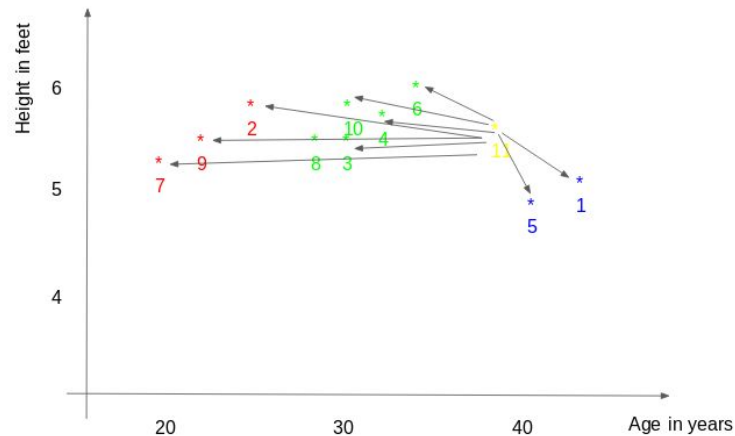
KNN Notes

- KNN is also known as k-nearest neighbors.
- The goal of KNN is to create groups of data based on how similar different data points are.
- This similarity is based on the distance formula.



How Does This Work?

- First, the distance between each point is calculated.
- The closest k points are selected.
- The average of these points is the prediction for the new point.



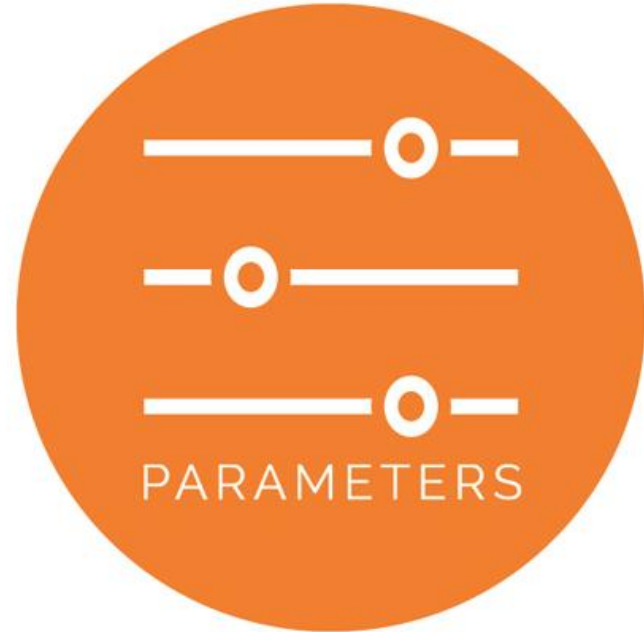
How Do You Calculate Distance?

- There are tons of formulas for calculating distance. These include:
 - Euclidean Distance
 - Manhattan Distance
 - Hamming Distance



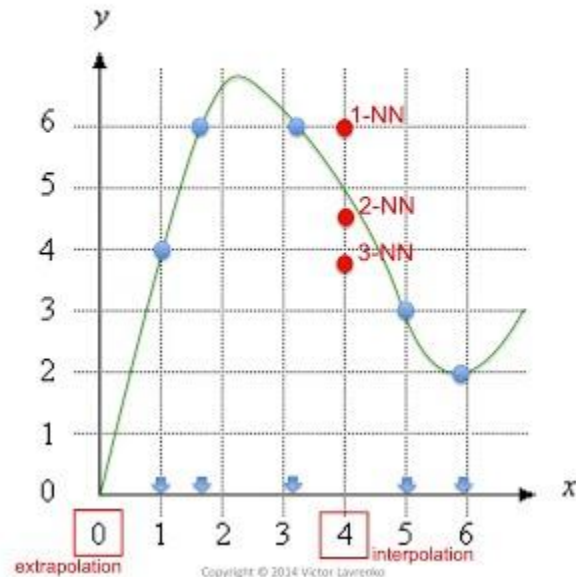
Parameters

- Unlike linear regression, KNN requires the use of parameters.
- The main parameters you will be using are k and distance formula.
- The ability of your model to make an accurate prediction directly depends on how well you choose these two parameters.



KNN Regression in Practice

Example: kNN regression in 1-d



Exit Ticket

Why might you want to use KNN regression instead of linear regression?