KNN Algorithm | LAB 9

What is Classification?

- Classification is a type of supervised learning. It specifies the class to which data elements belong to and is best used when the output has finite and discrete values. It predicts a class for an input variable as well.
- There are 2 types of Classification:

1. Binomial
2. Multi-Class

Classification use case

- To find whether an email received is a spam or not
- To identify customer segments
- To find if a bank loan is granted
- To identify if a kid will pass or fail in an examination

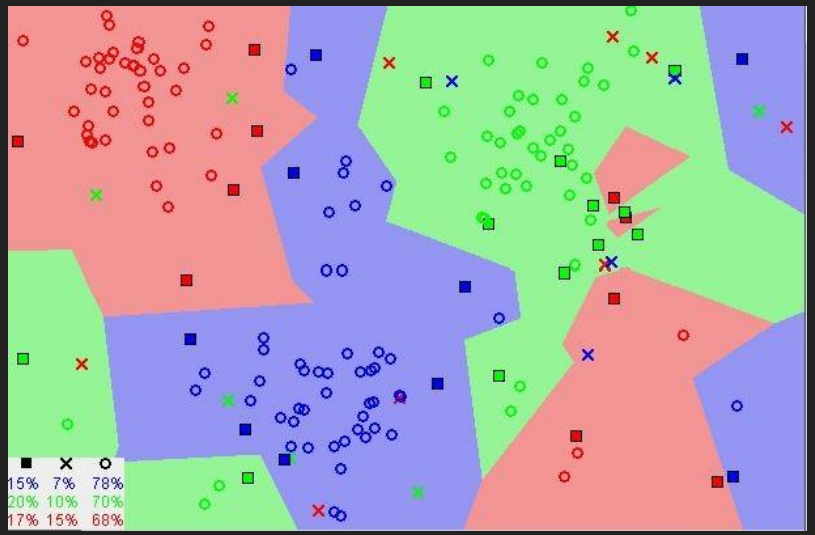
What is K-Nearest Neighbors?

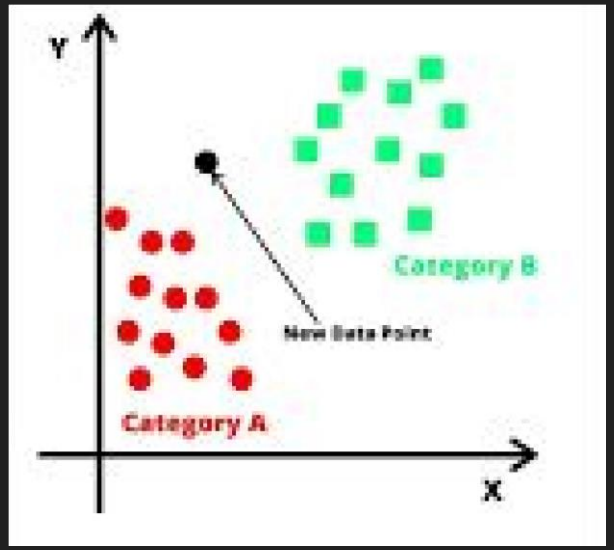
- The KNN algorithm assumes that similar things exist in close proximity. In other words, similar things are near to each other.

“Birds of a feather flock together”

- The KNN algorithm hinges on this assumption being true enough for the algorithm to be useful.
- It is used for classification problems.

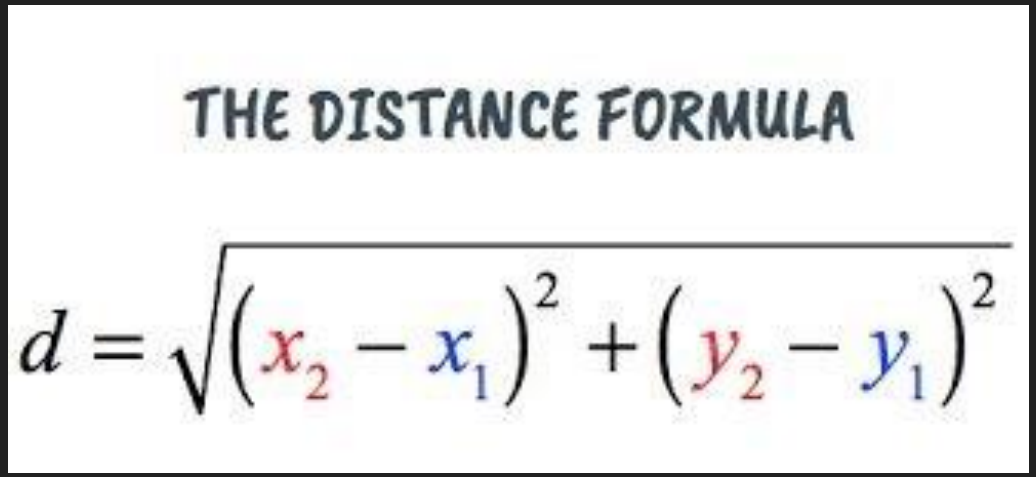
Example of KNN





Distance Formula

KNN captures the idea of similarity with some mathematics we might have learned in our childhood— calculating the distance between points on a graph.



The KNN Algorithm

1. Load the data
2. Initialize K to your chosen number of neighbors
3. for each example in the data
   - **3.1** Calculate the distance between the query example and the current example from the data
   - **3.2** Add the distance and the index of the example to an ordered collection
4. Sort the ordered collection of distances and indices from smallest to largest (in ascending order) by the distances
5. Pick the first K entries from the sorted collection
6. Get the labels of the selected K entries
7. If regression, return the mean of the K labels
8. If classification, return the mode of the K labels