K-Nearest Neighbour (KNN)

> KNN is one of the simplest algorithm based on Supervised Learning technique.

> K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.

> It stores all the available, data and classifies a new data point based on the similarity.

> K-NN algorithm can be used for Regression as well as for classification but mostly it is used for the classification problems.

> It is a non-parametric algorithm, which means it does not make any assumption on underlying data.

> It is also called a lazy learner algorithm because it does not learn from the training set, instead it stores the dataset. At the t

KNN algorithm.

1) Load the data
2 Initialize k to your choosen 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

How to select the value of K?

- There is no particular way to determine the best value for 'K', so we need to try some values to find the best out of them. The most preferred value for K is 5.
- 0

Advantages & Disadvantages

> It is simple to implement
> It is robust to the noisy training data.
> It can be more effective if the training data is large.

- Always needs to determine the value of K which may be complex some time.
- The computation cost is high because of calculating the distance between the data points for all the training samples.