K-Nearest Neighbors Algorithm

K-nearest algorithm is simple, supervised machine learning algorithm that can be used to solve both classification and regression problems. However, it is more widely used in classification problem in the industry. KNN is an approach to data classification that estimates how likely a data point is to be a member of one group or the other depending on what group the data points nearest to it are in. KNN algorithm can complete with most accurate models because it makes highly accurate predictions.

The k-nearest-neighbour is an example of a "lazy learner" algorithm, meaning that it does not build a model using the training set until a query of the data set is performed.

The advantage of KNN algorithm are:

1. Simple to implement
2. Flexible to feature/distance choices
3. Naturally handles multi-class cases
4. Can do well in practice with enough representative data.

Disadvantage of KNN algorithms are:

1. Need to determine the value of parameter K (number of nearest neighbors)
2. Computation cost is quite high because we need to compute the distance of each query instance to all training samples.
3. Storage of data.
4. Must know we have a meaningful distance function.