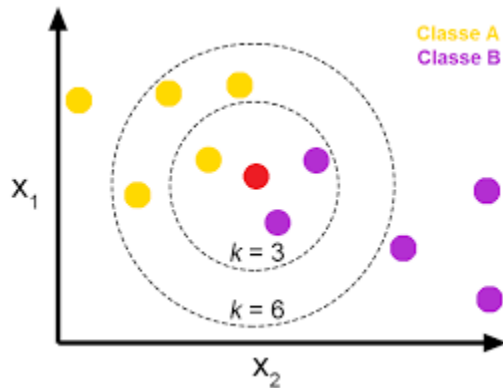


KNN Cheat Sheet

KNN (K Nearest Neighbour):

- KNN is simple Machine Learning Algorithm that comes under the supervised learning
- **KNN works on the assumption between the similarity between the new data points and available cases .**



- Case is classified by a majority vote of its neighbors data points . Assignment of case with respective to class is done by its K nearest neighbors measured by a distance function.

Distance functions:

1. Euclidean Distance

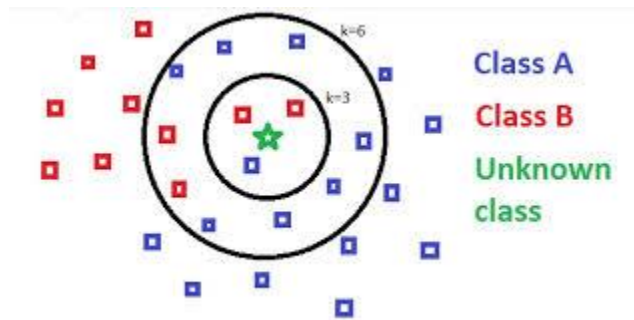
$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Manhattan Distance

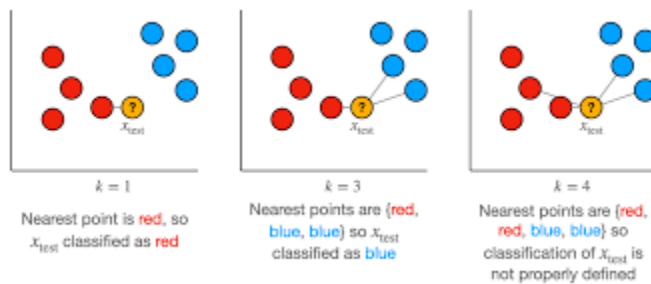
$$distance = \sum_{i=1}^n |p_i - q_i|$$

- KNN algorithm stores all the available data and classifies a new data point based on the similarity
- In case of the continuous features above mentioned distance techniques will be followed up however for categorical features the Hamming distance must be used

- It takes the decision based on the majority voting



- We can change its accuracy by changing the value of the perimeter named `n_neighbors`, its used as a hyper perimeter



- There are few approaches to get the value of K
 - 1) By using for loop
 - 2) By using Grid search cv
 - 3) Under root of total number of data pints

Advantages of KNN algorithm:

- Simple and easy to use
- KNN algorithm solves classification as well as regression problems
- New data can be added which doesn't impact the accuracy of the algorithm

Disadvantages of KNN algorithm:

- KNN algorithm doesn't fit with the large datasets
- Features standardization and normalization becomes important as we are dealing with distance measures before training KNN model
- KNN is pretty much sensitive to missing values and outliers

Applications:

- Recommendation system
- Credit Card Ratings Prediction

- **Loan Default Predictions**