K Nearest Neighbors

```
KNN:-
    1.KNN is supervised machine learning algorithm
    2.The algorithm can be used to solve both classification and regression problem
statements
    3.It is easy to implement and understand
    4.It is non - Parametric algorithm (No assumption on data)
    5.Lazy learner algorithm
    6.It is distance based algorithm
```

```
KNN is Distance based algorithm
There are two types of distance:-1.Euclidian distance
                                 2.manhattan distance

1.Euclidean Distance:-It is a measure of the true straight line distance between two
points in Euclidean space.

2.Manhattan Distance:-Manhattan distance is sum of all the real distances between
source(s) and destination(d) and each distance
                      are always the straight lines.
```

KNN Algorithm:

```
1. We need to select the value of K (Number of Neighbors)
2. Calculate distance between new datapoint (testing datapoint) with all the remaining
data points
   (training datapoints)

K = 5 (Default value of K)

K < 5 >> It can be noisy, Low bias and high variance
```

Featuring Scaling

```
1. MinMax Scaler (Normalization)
2. Standard Scaler (Standardization) - Z Score
3. Robust Scaler
4. Unit Vector Scaler
5. Power Transform

Most of we use only two scaler in machine learning.

MinMax Scalar(Normalization):-
    1.Minimum and maximum value of features are used for scaling
    2.It is really affected by outliers
    3.It is useful when we don’t know about the distribution
    4.It is a often called as Scaling Normalization
    5.Scales values between [0, 1] or [-1, 1].
```

```
    6.Xnormalization = (X - Xmin) / (Xmax - Xmin)

Standard Scaler(Standardization):-
    1.Mean and standard deviation is used for scaling
    2.It is much less affected by outliers.
    3.It is useful when the feature distribution is Normal or Gaussian
    4.It is a often called as Z-Score Normalization
    5.It is not bounded to a certain range.

    6.Xnew = (X - Xmean) / std

         z = (x - xmean) / Std
```

KNN Classification

```
1. Select the value of k
2. Find out dist between new data point to all remaining data points
3. Select the nearest K distances (Data points)
4. Use voting classifier (5B and 3R) >> B
```

KNN Regression

```
1. Select the value of k
2. Find out dist between new data point to all remaining data points
3. Select the nearest K distances (Data points)
4. Use mean / Average of K neighbors
```

Advantages

```
1. Easy to understand and easy to implement
2. Non - Parametric (No - Assumptions on the data)
3. Can be used for Regression as well as classification
4. No training steps
5. It naturally lends itself to multiclass classification
6. Only one hyperparameter to be set (k-value, p-value)
```

Disadvantages

```
1. Lazy Learner (Testing stage is very slow)
2. Need to do feature scaling
3. Every time we need to select the value of K
4. Sensitive to outliers
5. Can not use KNN for high - Dimensional data (10000, 20000)
   (Dimensionality Reduction technique) - PCA
```

Hyperparameter Tuning

In [ ]:

```
Hyperparameter Tuning:-it is the process of choosing the optimum set of hyperparameter for
                      is also called hyperparameter optimization.
There are two types of parameters
1.Model Parameter:-These are the parameters of the model that can be determined by traning
                 considered as internal parameters.

2.Hyperparameters:-Hyperarameters are parameters whose values control the learning process.
                 to obtain an optimal model.External parameters.

    We need to search for right set of hyperparameters
    Hyperparameters ( k- Value, p- value)

k_value = [1 to 20]
p_value = [1,2]

1. GridsearchCV :
    It will try every combination of present list parameters value in accurate way.

2. RandomizedSearchCV :
    Random combination
    It goes through only fixed number of combination
```