

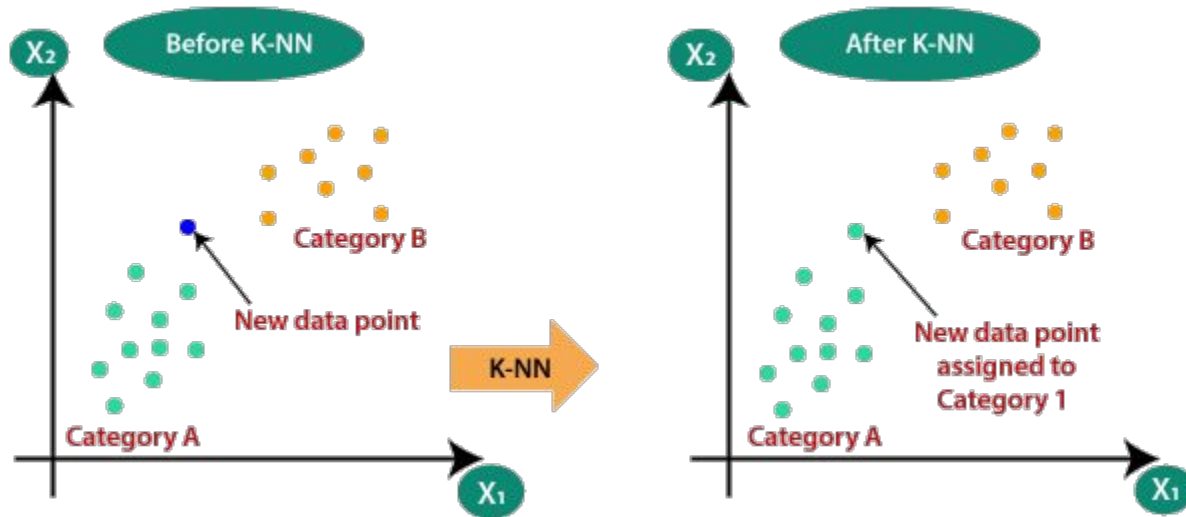
### 3) KNN (K-nearest neighbours) :

--> Knn algorithm is a simple,easy-to-implement,lazy learning, Supervised machine learning algorithm

--> This can be solved using both classification and regression problems using feature similarity.

--> K is number used to identify similar neighbours for the new data points.

--> KNN takes K nearest neighbors to decide where the new data point with belong to.This decision is based on feature similarity.



## Steps Involved In KNN:

Step:1) choosing the value for K, where K value should be in Odd number.

Step:2) Finding d/s of the newpoint to each of the training dataset.

Step:3) Finding the K nearest neighbors to the new data point.

Step:4) For Classification: Count the number of data points in each category among the K neighbours. New datapoint will belong to the class that has the most neighbours.

For Regression: Value for the new datapoint will be the average of the k Neighbours.

KNN distance is calculated by:

1) Euclidean Distance.

2) Manhattan Distance.

#### Distance functions

Euclidean

$$\sqrt{\sum_{i=1}^k (x_i - y_i)^2}$$

Manhattan

$$\sum_{i=1}^k |x_i - y_i|$$

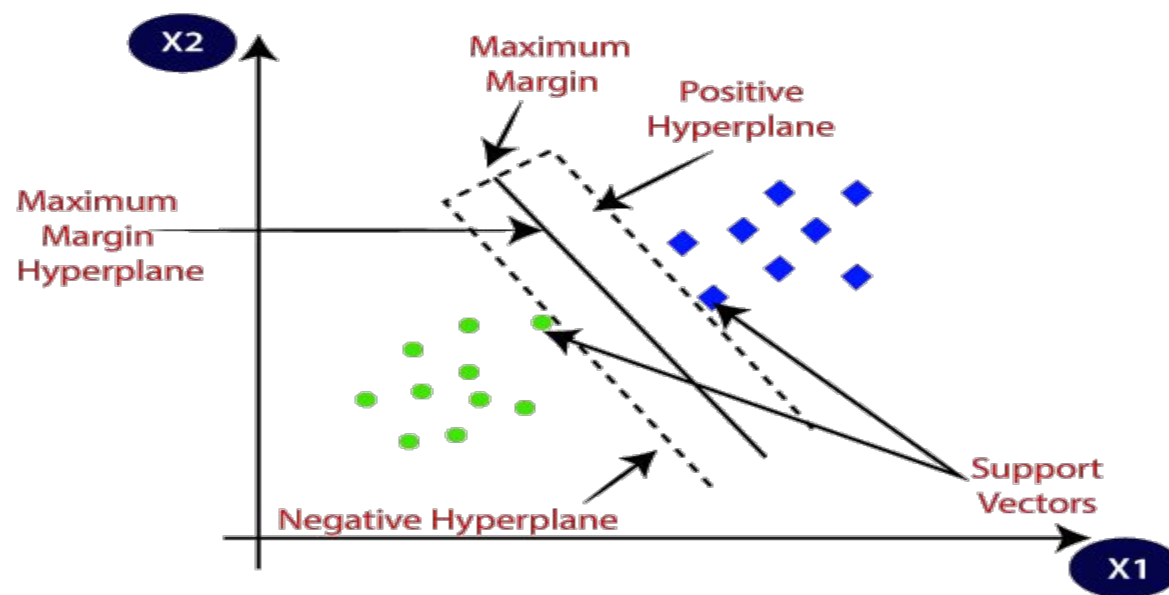
## Applications:

- 1)Text mining.
- 2)Agriculture.
- 3)Finance.
- 4)Medical.
- 5)Facial recognition.
- 6)Recommendation systems

## 4) Support Vector Machine (SVM) :

--> It is the most popular ML Algorithm, which is used for both Classification and Regression problems.

--> SVM algorithm helps to find the best line or decision boundary, called hyperplane that classify the data points.



## SVM types:

- 1) Linear SVM(LSVM): Used for separable data, which means if a dataset can be classified into two classes by using single straight line, then such data is termed as Linearly separable data.
- 2) Non-linear SVM(NLSVM): Used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data.

## Hyperplane:

--> The SVM algorithm helps to find the best line or decision boundary, this best boundary or region is called as a hyperplane.

--> We always create a hyperplane that has a maximum margin, which means the maximum distance between the data points. The hyperplane with maximum margin is called the optimal hyperplane.

## Support Vectors:

--> The data points or vectors that are the closest to the hyperplane. SVM algorithm finds the closest point of the lines from both the classes. These points are called support vectors.

**Margin:** The distance between the support vectors is called as margin and the goal of SVM is to maximize this margin.

## Kernel Trick / Kernel function:

A kernel transforms a low-dimensional input data space into a higher dimensional space. So, it converts non-linear separable problems to linear separable problems by adding more dimensions to it.

Thus, the kernel trick helps us to build a more accurate classifier. Hence, it is useful in non-linear separation problems.

In the context of SVMs, there are 4 popular kernels - Linear kernel, Polynomial kernel, Radial Basis Function (RBF) kernel and Sigmoid kernel.

#### Linear Kernel:

->It helps to draw a best fit line that separates both the classes linearly.

->Linear kernel support vector machines have good performance only on very simple problems.

#### Polynomial Kernel:

->In the polynomial kernel, we simply calculate the dot product by increasing the power of the kernel.

#### Radial Basis Function(RBF):

->RBF is the most popular support vector machine kernel choice, and the default one used in sklearn.

->RBF is short for "radial basis function". It is a general-purpose kernel; used when there is no prior knowledge about the data.



## Sigmoid Kernel:

->Mostly, it is used in deep learning, to solve ANN based problems.

## C and Gamma in SVM:

->Since kernels transforms data into higher-dimensional spaces , machine learning algorithms making use of the kernel trick have a tendency to overfit.

->The sklearn SVM methods provide C and gamma functions for this purpose.

->C & Gamma is a hypermeter which is set before the training model and used to control error .

## Applications:

1)Face Detection.

2)Text and hypertext categorization.