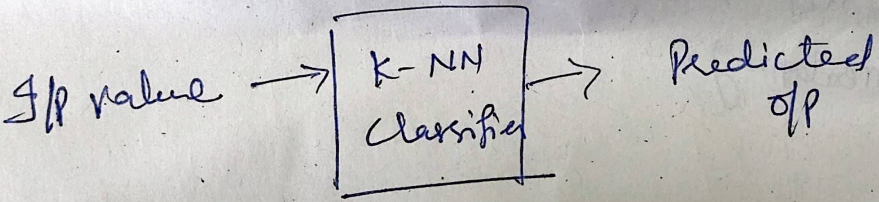
K-NN → (K- nearest Neighbours)

Algorithm

KNN is one of the simplest supervised machine learning algorithms used for Classifications.

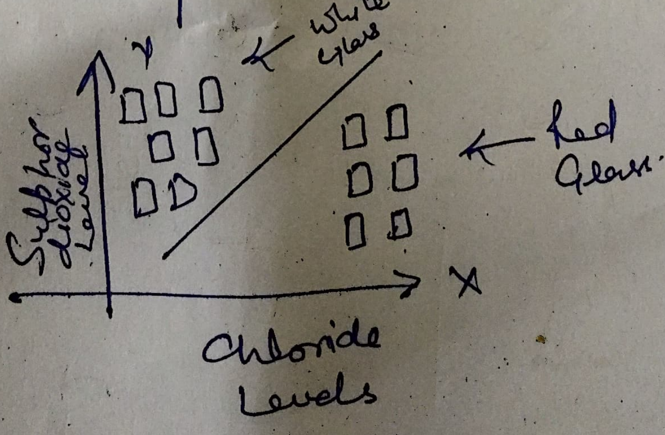
- It classify data points based on it neighbouts classifications. It stores all available cases and classifies new cases based on similar features.

- K-NN is based on feature similarity, we can perform classification task, using K-N-N classifies algorithm.



Example :- Prediction of glass of a ~~wine~~ Ball is red/ white.

Different Variables are considered in this KNN algorithm, including Sulphor dioxide & chloride Levels.



K- in KNN is a parameter that refers to the number of nearest neighbours in the majority voting process.

Here, if we take K=5, the majority votes from its fifth nearest neighbor and Classifies the data point. The glass of wine Will be classified as red since four out of five neighbours are red.

How to choose the factor - K

- Selecting the right K value is a process called parameter tuning, which is important to achieve high accuracy.
- There is no defined way to choose the value of K, it depends upon the type of Problem using are Solving.
- We try Some Values to find the best out of them. The most Preferred Value for K ~~is~~ 5.
- A very low value for K such as K=1 or K=2, can be noisy & Leads to the effect of outliers in the model.
- Large value for K are good, but it may find some difficulties.

KNN – is a non Parametric algorithm, Which means it does not make any assumptions on underlying data

– KNN is also called a Lazy Learner Algorithm, because it does not learn from the training set immediately, instead it stores the data set, and at the time of classification, it performs an action on the data set.

– K-NN algo. at the training phase, just stores the data set and when it gets new data, then it classifies that data into a category that is much similar to the new data.

Why ~~We~~ do we need K-NN Algo

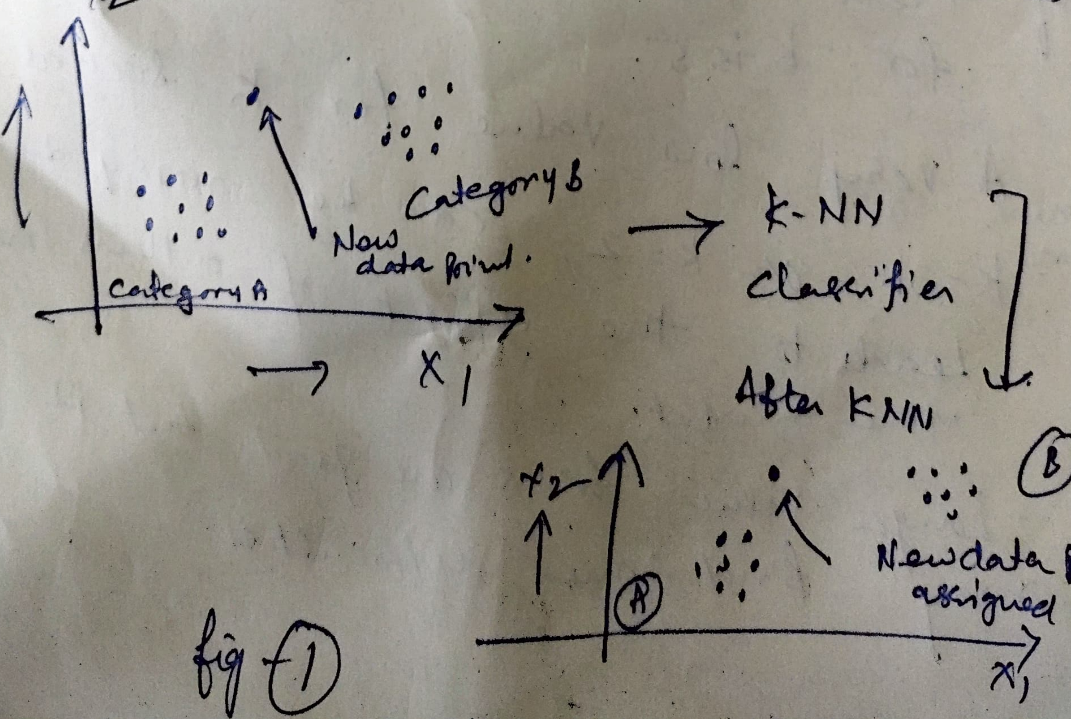


fig (1)

How does K-NN work

Step-1 – Select the number K of the neighbours.

Step-2 – Calculate the Euclidean distance of K number of neighbours.

Step-3 – Take the K-nearest neighbours as per the calculated Euclidean distance

Step-4 – Among these K-neighbours, Count the number of data points in each category

Step-5 – Assign the new data points to that category for which the number of the neighbour is maximum.

Step-6 – our model is ready.

Example :- Let take an example of fig)

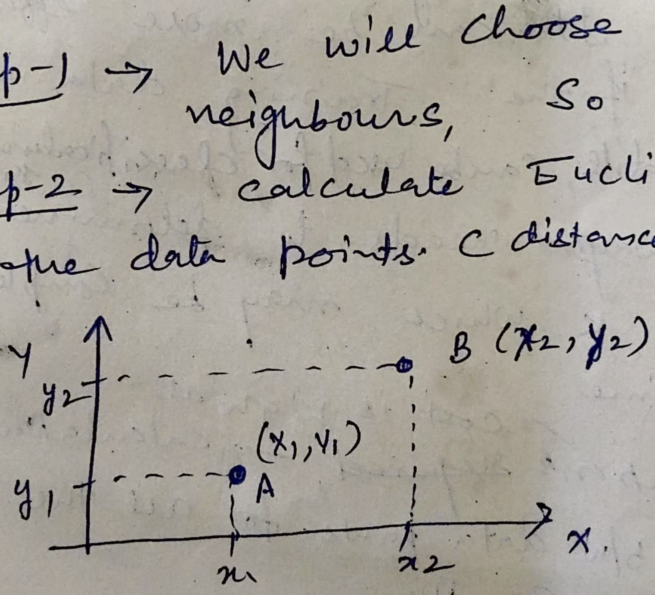
We have a new data point and we need to put it in the required catogory.

✓ **Step-1** → We will choose the number of neighbours, So we choose K=5

✓ **Step-2** → calculate Euclidean distance b/w the data points. (distance b/w two data points)

EU b/w A, & B,

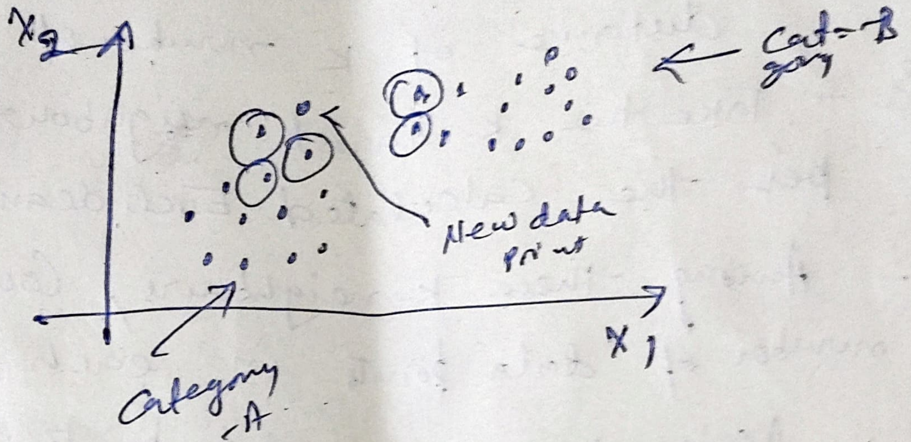
$$\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$$



Step-3 – By calculating the Euclidean distance, We got the nearest neighbours,

– 3- Nearest neighbours in Cat-A

2- Nearest neighours in cat b.



As 3- nearest neighbours are from category A, hence this new data point must belong to Category A.

Advantages :– 1. It is Simple to implement

2. It is robust to the noisy training data

3. It can be more effective if the Training data is large.

4. Algo ~~it~~ is versatile, canbe used for classification/Regre task

Disadv → 1. Always needs to determine the Value of K which may be Complex some time.

2. The Computations ~~required~~ cost is high as calculation of distance b/w data points for all the training samples.