KNN - K-Nearet Neighbor.
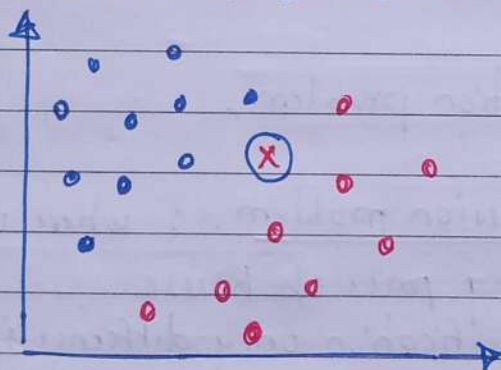
• The abbrevation KNN stands for "k- Nearet Neighbor" it is supervised machine learning algorithm this algorithm can be used to solve both classification as well as regression problem.

• The number of nearest neighbor to a new unknown Variable that has to be predicted or classified denoted by the symbol 'k'

• whenever new data will come if new data is close to 1 then prediction will be class 1 other wise 0 for binary classification same principle for multiclasification problem.

• In general K is odd number

Let understand with some example

Now we have to find new added ⊗

Working of KNN Algorithm.

step 1: ladding training as wellas teet data.

step 2: Next we choose value of K (hyperparameter) the nearest datapoints. K can be any integer.

step 3: for each data point (new) teal dota do the following.

3.1> calculate distance between teat data and each row of training data with the help of distance calculating method. like Euclidion, Manhatten or Hamming distance. The most commonly method to calculate is Euclidion.
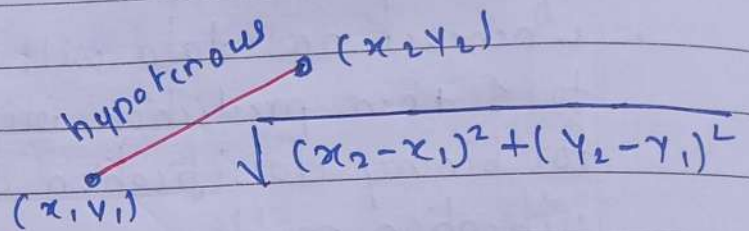
3.2) Now based on its distance value, sort them in ascending order.

3-3) Next if will choose the top k rows from sorted array.

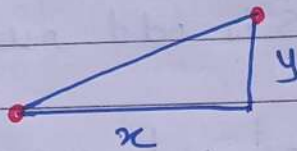3.4) Now it will assign a class to a test point based on most frequent class of these rows.

step 4 : End.

a) Eucledian distance :-

hypotenous

(x_2, y_2)

(x_1, y_1)

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

b) Manhattan distance

y

x

M. distance $= x + y$.

this is for classification problem.

Now let see for regression problem. : where output is continuous data. e.g price of house.

1). it is similar to classification only difference in last step where we were considering most frequent class here from k number of output (nearest 'k') we calculate their mean, that's it

Limitation : Not applicable to huge dataset since calculating distance would consume lot of time

• Senshive to outliers

• Senshive to missing values