Date

Expt. No.

Page No.

① $\sqrt{(6-4)^2 + (8-3)^2} = \sqrt{29} = 5.38$

② $\sqrt{(6-6)^2 + (8-7)^2} = (1)$ - Pass

③ $\sqrt{(6-7)^2 + (8-8)^2} = (1)$ - Pass

see the nearest value

④ $\sqrt{(6-5)^2 + (8-5)^2} = \sqrt{10} = 3.16$

⑤ $\sqrt{(6-8)^2 + (8-8)^2} = (2)$ - Pass

for given x → as K=3 we have to consider 3 value

⇒ P P P = 3 [Based on nearest neighbour]

fail = (0)

3 > 0

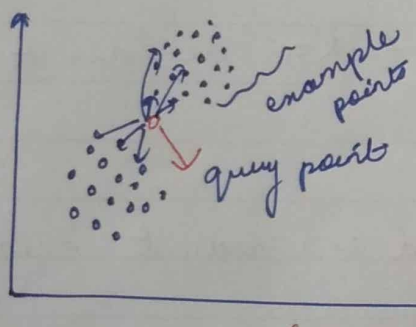
So x is declare to be Pass

* KNN → It is a supervised classification learning algorithm is which you have some data points or data vector which is separated into different no of several categories or classes and it try to predict the classification of a new sample from a particular population set

Teacher's Signature

⇒ A lazy algo^m → (It this to only memorizes the process . It does not learn any process)
(means it does not take it own decision suppose you call a KNN algorithm to go there it will go there you will call it to come here it will come to that point

⇒ It classifies new points based on a similarity measure like - euclidean distance



Assume N = 30 datapoints

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

Suppose we choose any point and we find the neighbourhood of that point
It find out the distance between query point and example point with help of euclidean distance or Minkowski distance

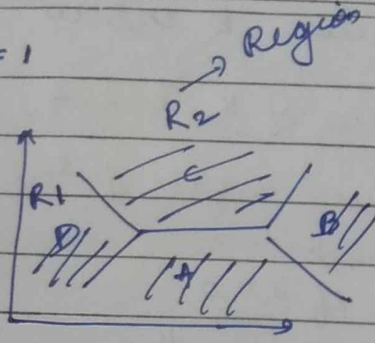
* - K must be odd eg (3, 5, 7)
→ K must not be a multiple of classes
if K is even then there will be tie in the classification of that particular label

eg if C=7 K ≠ 14

Date

Expt. No. Page No.

if K = 1



Then the particular data point will be assigned certain region into the partition

Basically it partition the space into n no of region. This is called voronoi partition space

It means if there is any data element falling in R_1 will be assigned to D class label

similarly $R_2 \rightarrow C$ class label

Algo?

S1: load Data from csv or xls file

S2: Initialize K ⇒ hyperparameter [Assign nearest neighbour to it (any odd no)]

S3: for each sample in training data.

S31. calculate distance between query point & the current point

S32 - Add the distance & the index of the example to an ordered collection

S4: Sort ⇒ ordered collection of distance & indexes from small to large (ascending order)

Teacher's Signature

S_5 - Pick first K entries from sorted collection

S_6 - Get the labels of selected K entries

if regression → return mean of K labels

if classification → return mode of K labels

ex

if K = 3

mode = K

Advantages of KNN algo"

→ It is simple to implement

→ It is robust to the noisy training data

→ It can be more effective if the training data is large

Disadvantages of KNN algo"

→ Always needs to determine the value of K which may be complex some time.

→ The computation cost is high because of calculating the distance between the data points for all the training sample

* How to select the value of K in the K-NN Algo"?

Below are some points to remember while selecting the value of K in the K-NN algo".

Date _____

Expt. No. _____ Page No. _____

→ There is no particular way to determine the best value for "K" so we need to 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 and lead to the effects of outliers in the model.

→ Large values for K are good, but it may find some difficulties.

✱ Support vector Machine

It is one of the most popular supervised learning algo^n which is used for classification as well as Regression problem. The goal of SVM algo^n is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in future.

⇒ The best decision boundary is called a hyperplane

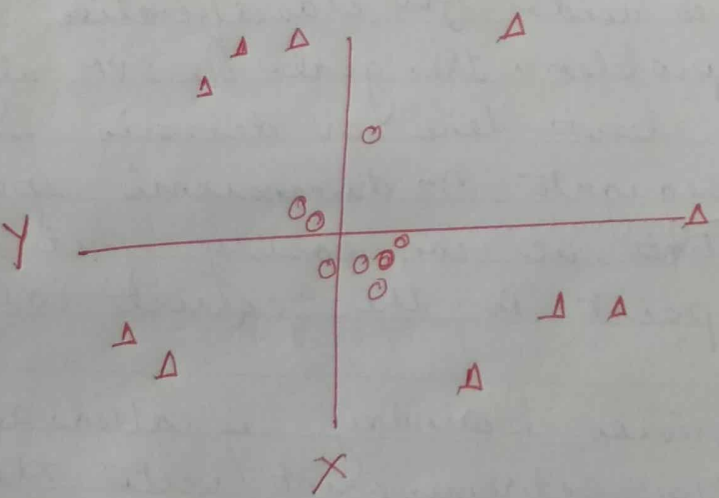
⇒ SVM chooses the extreme point/vector the helps in creating the hyperplane. These extreme cases are called as support vector. & hence the algo^n is termed as SVM

Teacher's Signature _____

Types of SVM

① Linear SVM - Linear SVM is used for linearly separable data which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data & classifier is used called as linear SVM classifier.

② Non-linear SVM - Non-linear SVM is 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 and classifier used is called as Non-linear SVM classifier.



Non-linear SVM

Date

Expt. No.

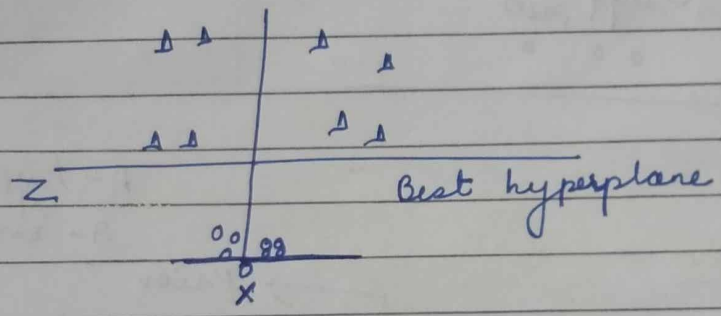
Page No.

* if data is linearly arranged, then we can separate by using a straight line, but for non-linear data, we cannot draw a single straight line

→ So to separate these data points, we need to add one new dimension

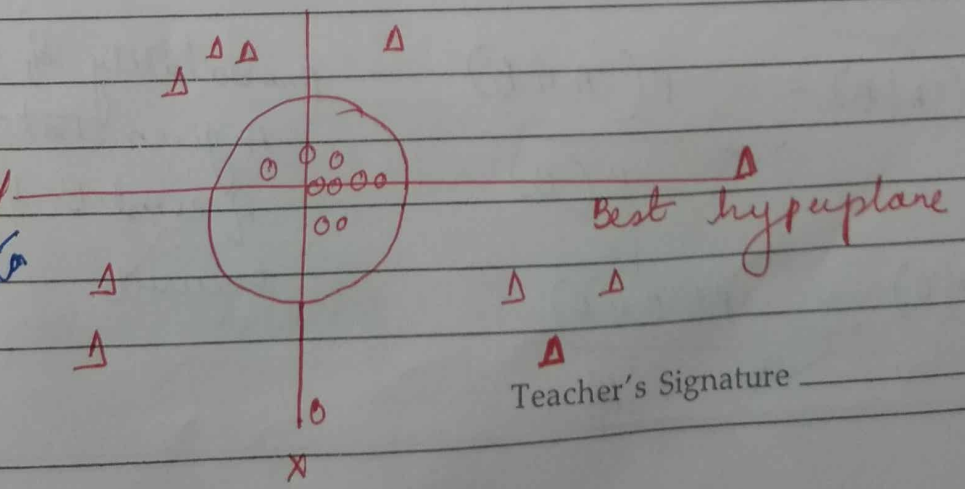
$$x, y, z$$

$$z = x^2 + y^2$$



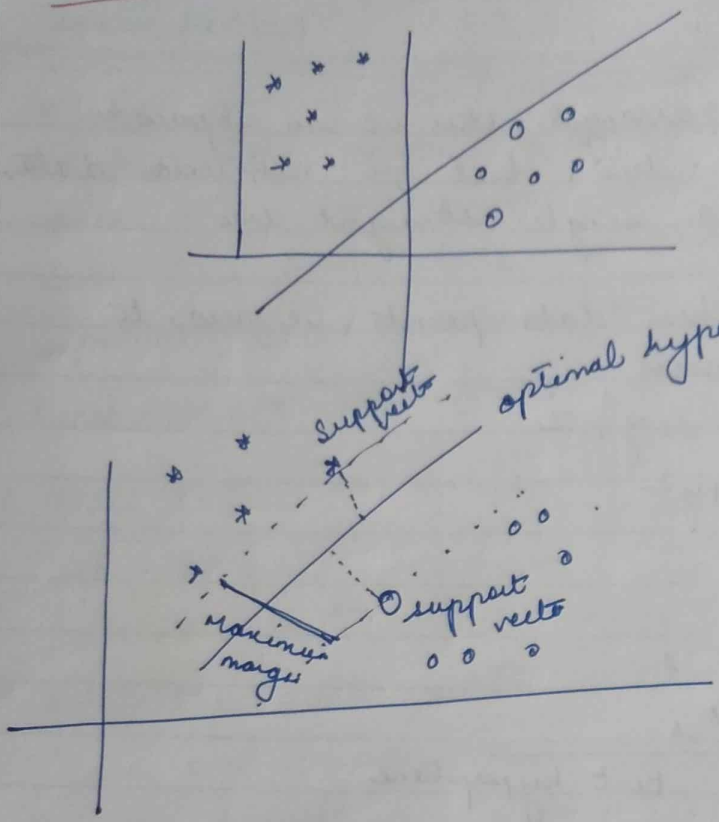
Since we are in 3-d space, hence it is looking like a plane parallel to the x-axis.
If we convert it in 2d space with z=1, then it will become a

we get a circumference of radius 1 in case of non-linear data



Teacher's Signature

Linear SVM



A - hypothesis
B - evidence / data

Bayes Theorem

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

- $P(A|B)$ → Posterior
- $P(B|A)$ → likelihood
- $P(A)$ → Prior
- $P(B)$ → marginal

- $P(A|B) = \frac{P(A \cap B)}{P(B)}$

Probability of A event A given that probability of event B has already occured

- $P(B|A) = \frac{P(B \cap A)}{P(A)}$