Date .....

Support Vector Machine

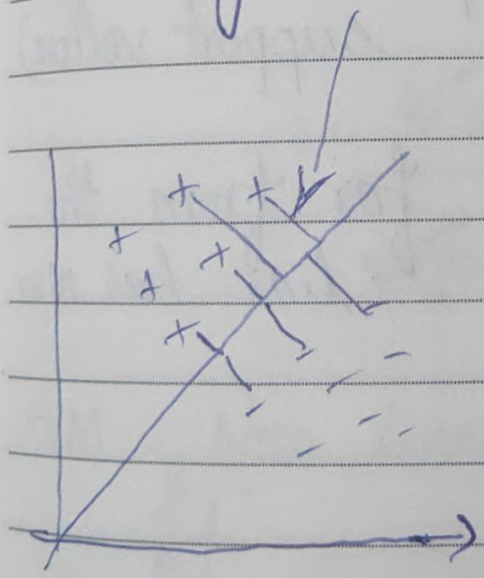
→ Supervised ML Algorithm. (Mainly used for)

→ Used for both (classification) & Regression.

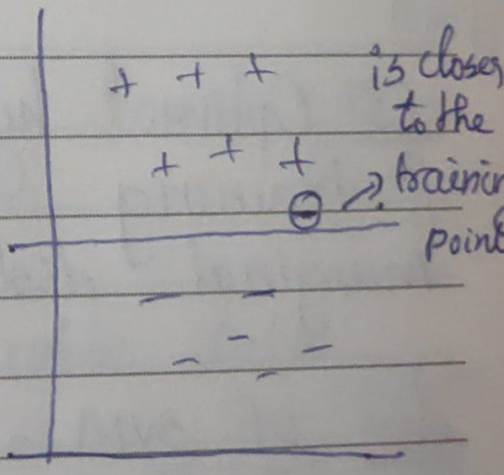
→ Non-Probablistic classifier.

→ Both linearly & non-linearly seperable data

→ Marginal distance



(A) ✓

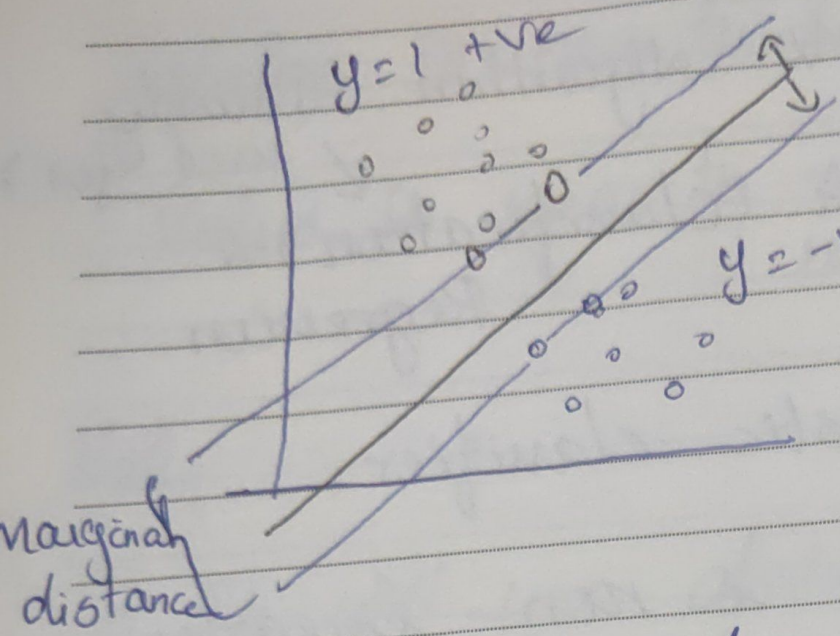


(B)

Note:-

1D → Threshold
2D → line
3D → Plane
4D → hyperplane
D → Dimension.

SVM - Linearly Separable data



(1) Support vectors.

(2) Marginal lines (passes through support vectors)

(3) Optimal hyperplane: far from the training data pts. (line has max. marginal distance)

Aim of SVM:-

To get an optimal hyperplane which has maximum marginal distance

Date .....

Eqn of Maginal line in -ve side

$$W^T x + b = -1$$

Eqn of Maginal line in +ve side

$$W^T x + b = 1$$

Eqn of hyperplane $W^T x + b = 0$ [Assume the plane passes through origin]

Distance - $W^T x_1 + b = 1$

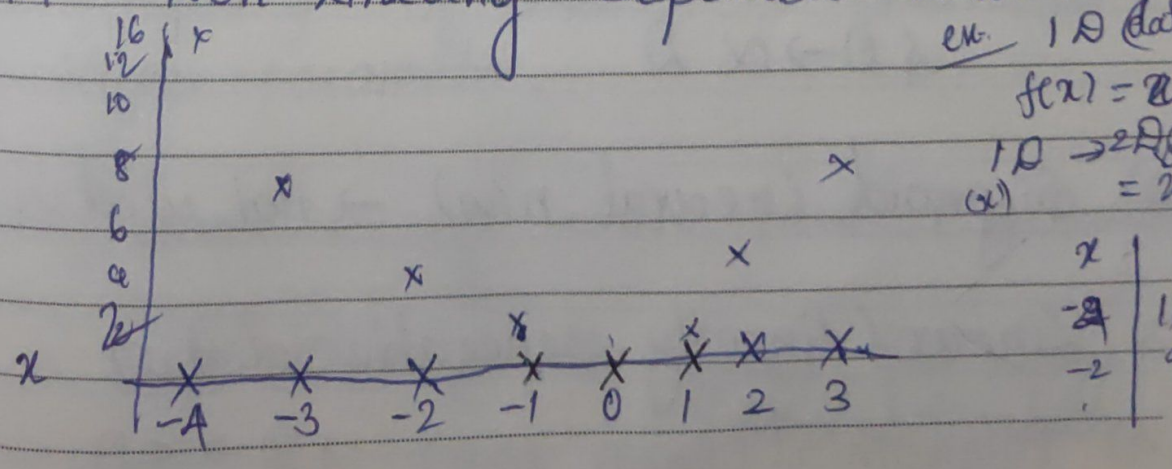
$\rightarrow W^T x_2 + b = -1$

$W^T(x_1 - x_2) = 2$

$\Rightarrow x_1 - x_2 = \frac{2}{\|W\|}$

$\text{Max} = \frac{2}{\|W\|}$

SVM - Non Linearly Seperable data



ex. 10 (data)

$f(x) = x^2$

$10 \rightarrow 2(x) = x^2$

x	y
-4	16
-2	4

Date .....

→ SVM uses a set of mathematical function (kernel function)

Kernel function - Used to convert the lower dimensional data to a higher dimensional data & find the linearity.

→ finds the rltnshp in higher dimension using kernel function.

→ Transforms the data into required form.

Kernel function:-

1) RBF - Radial Basis function (creates infinite dimensions)

4D → ∝ D

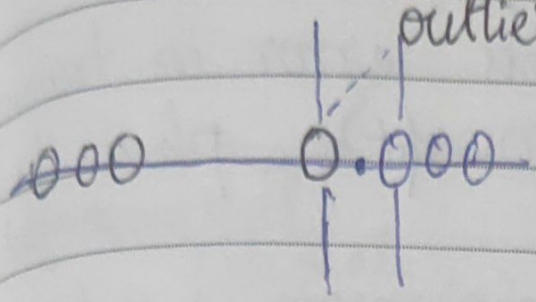
2) Sigmoid (neural n/w) → not used in SVM

3) Linear (Linearly seperable data)

4) Polynomial (linear eqn) 4D → 5D
2D → 3D.

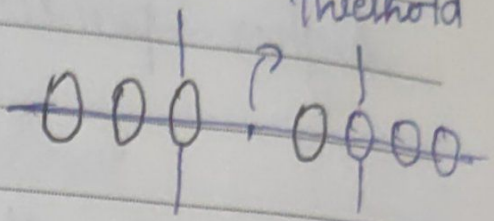
Date .....

Hard Margin Classifier



→ Doesn't accept errors in the training phase, We will have more errors in testing phase.
(ie.) Overfit.

Soft Margin Classifier.



→ accepts error in the training phase. (to have generalized result). to have less error in the testing phase.

Hyper parameter:-

→ Defines the architecture of a model

→ Training of model is controlled by Hyper parameter.

→ Hyper parameter of SVM used 'C' & 'Gamma'

→ C → The penalty given for every misclassification.

Date .....

creating the model → value - C.

higher penalty = less error in training phase (hard margin) (overfit)

Lower Penalty = more error → Underfit.

→ Optimal value for C → Use hype parameter tuning.

⊛ Hyper Parameter Tuning will be done to find optimal value for hyperparameter to improve the model performance.

Gamma :-

→ define the curvature of hyperplane

Let

Gamma = 0.1, 0.01, 0.001, 0.0001

C = 0.1, 0.01, 0.001, 0.0001.

→ which combination will give better score for ex. (Gamma = 0.001, C = 0.01) is the better score.

Date .....

Cross - Validation (CV):-

It takes entire data for Training & Testing. In CV score if std < 0.05 then the model is good.

Types

1) Leave One Out CV (LOOCV)

This method gives you Low bias & high variance.

Bias → errors for Training data
Variance → errors for new data.

→ Drawback of LOOCV

→ Overfitting
→ High computational process.
→ Time consuming.

2) K-Folds.

No of folds. Max : K = 10
Usually 5 or 10

k-fold Only for.

Date .....

3) Stratified (classification)

Balanced data while doing the split.

Hyper Parameter Tuning Technique

① Grid Seaech CV

② Randomized Seach CV.

① GRid Seach CV.

C = [0.1, 0.01, 0.001]

G = [0.1, 0.01, 0.001]

→ Try with all the combinations & find the combinations which gives good score.

② Randomized Search CV.

Randomly picks some of combinations & checks which combinations give good score.