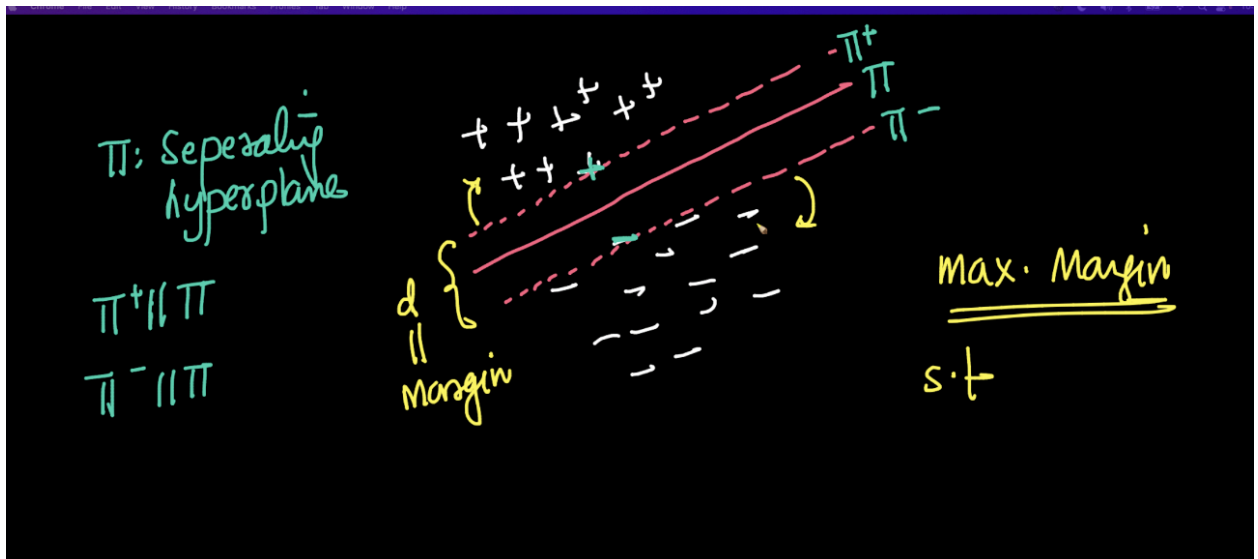


SVM-Introduction

- Support Vector Machine (SVM) is a powerful supervised machine learning algorithm.
- It works by finding the optimal hyperplane that best separates data points into different classes while maximizing the margin between the classes, making it highly effective in high-dimensional spaces.

What is the key idea behind SVM?

- The best hyperplane $\pi: w^T x + b = 0$ classifying between 2 classes is the one that has maximum gap/margin (d) between itself and the closest +ve and -ve datapoints.



Margin/gap

The hyperplane parallel to π , that touches the closest +ve point:

$$\pi^+: w^T x + b = 1;$$

The hyperplane parallel to π , that touches the closest -ve point:

$$\pi^-: w^T x + b = -1$$

Margin is measured as the distance between them: $d(\pi^+, \pi^-) = \frac{2}{\|w\|}$;

- where w is the weight of the model.

The optimization problem in SVM

Optimization problem: $\max \frac{2}{||w||}$

The goal is to maximize generalization.

Support Vectors

These are the data points:-

- Are within the margin
- Or, are misclassified
- Or, which lies on the hyperplanes (π^+ , π^-)
- $\alpha_i = 0$ for nonsupport vectors, whereas, $\alpha_i > 0$ for support vectors.

Different Types of SVM model

A. Hard Margin SVM:

- The simplest form of the SVM model.
- It assumes no data point can lie inside the Margin.
- Rarely works in real-life problems.

Hard Margin SVM

Our goal is to maximize the margin
s.t. the +ve and -ve samples are on different sides of hyperplane x

Therefore, the margin constraints are :

For $y_i = +1 \rightarrow x^* = y_i (w^T x + b) \geq 1$
For $y_i = -1 \rightarrow x = y_i (w^T x + b) \leq -1$

If we club both these equations:

$$\operatorname{argmax}_{w,b} \left(\frac{2}{||w||} \right) \text{ s.t. } y_i (w^T x_i + b) \geq 1$$
$$\forall i : 1 \rightarrow N$$

B. Soft Margin SVM:

- Introduces ζ as error for an incorrectly placed data point.

- $\zeta = 0$; datapoint is placed such that the hyperplane classifies the point correctly
- $\zeta > 0$; datapoint is placed such that the hyperplane classifies the point incorrectly
- $\zeta < 1$; datapoint is placed such that the hyperplane still manages to classify the point correctly, but lies inside the margin.
- Linear soft margin SVM is similar to LogReg

In general,

$$y_i(w^T x_i + b) \geq 1 - \zeta_i$$

Note:

- $\zeta_i = 0$ for all correctly placed points
 - $\zeta_i > 0$ for all misclassified points
- $\zeta_i \geq 0$ for all data points

Now, the optimization fn changes to :

$$\min_{w,b} \frac{\|w\|^2}{2} + \frac{C}{N} \sum_{i=1}^N \zeta_i \quad \text{s.t.} \quad (w^T x_i + b)y_i \geq 1 - \zeta_i$$

$$\forall i : 1 \rightarrow N \text{ and } \zeta_i \geq 0$$

This is known as Soft Margin SVM