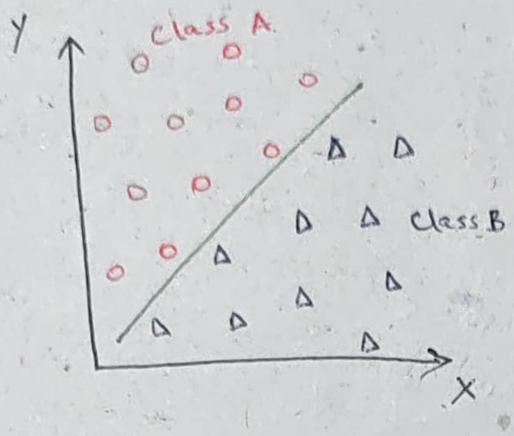
5. Support Vector Classifier (SVC)

Introduction :

Classification is an important tool in ML, which determines, "to which class an observation belongs".



→ Binary classification (2 classes).

Applications :

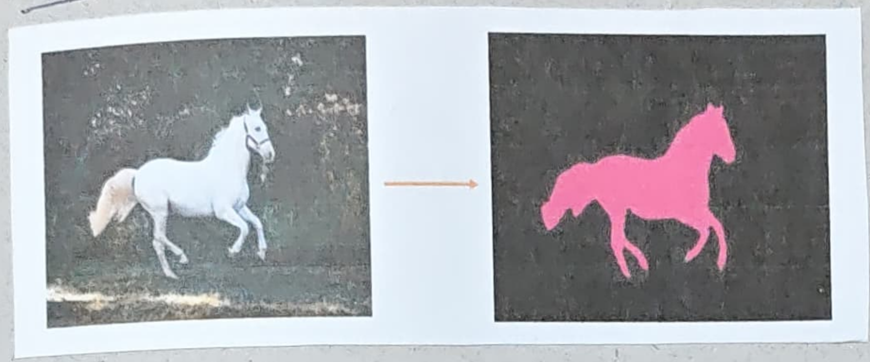


Image Segmentation :

- Classify pixels as belonging to foreground or background.

Linear classifier

⊙ Linear classifier corresponds to a hyperplane ($\bar{a}^T\bar{x} = b$) in N-dimensions.

⊙ Easy to determine and analyse!

⊙ General structure is

$$C_0 : \bar{a}^T\bar{x} \geq b$$

$$C_1 : \bar{a}^T\bar{x} \leq b.$$

$\bar{a}^T\bar{x}$ → Inner product b/w the vectors $\bar{a}$ and $\bar{x}$

Eg.

$$2x_1 + 3x_2 = 5$$

$$\Rightarrow \begin{bmatrix} 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 5$$

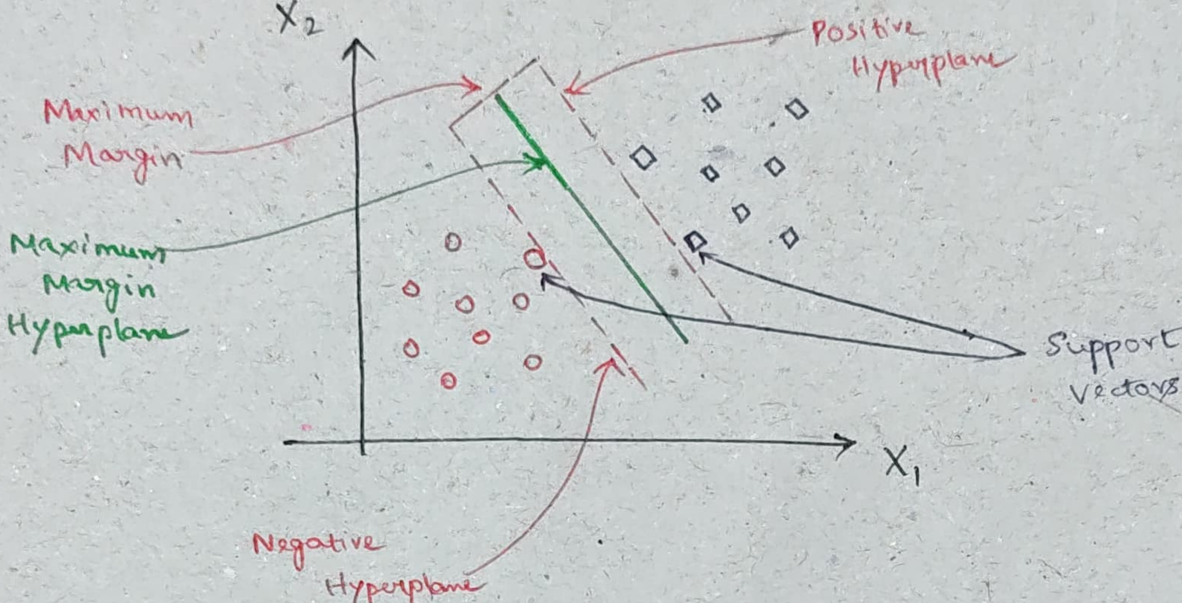
- How to determine the best Linear classifier ?

- consider the training set

$$\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_M \in C_0$$

$$\bar{x}_{M+1}, \bar{x}_{M+2}, \ldots, \bar{x}_{2M} \in C_1$$

- SVC has two parallel hyperplanes, which separates both classes by a slab.



- The classifier structure is

$$\bar{a}^T \bar{x}_i + b \geq 1, \quad \bar{x}_i \in C_0$$

$$\bar{a}^T \bar{x}_i + b \leq -1, \quad \bar{x}_i \in C_1$$

- The width of the slab is termed as Margin.
- Best classifier maximizes the margin between the two classes.

- Distance between the Hyperplanes.

- consider the hyperplanes

$$\left.\begin{aligned} \bar{a}^T \bar{x} &= c_1 \\ \bar{a}^T \bar{x} &= c_2 \end{aligned}\right\}$$ Parallel planes. Hence slopes $(y = mx + c)$ are equal

- Distance between the hyperplanes is

$$\frac{|c_1 - c_2|}{\|\bar{a}\|}$$

$|\cdot| \rightarrow$ Distance should always be +ve.

$\|\cdot\| \rightarrow$ Norm.

Problem 1 : What is the distance between the Hyperplanes

$$x_1 + 2x_2 + 3x_3 + \dots + N x_N = 1$$

$$x_1 + 2x_2 + 3x_3 + \dots + N x_N = -1$$

$c_1 = 1$, $c_2 = -1$ and $\bar{a} = \begin{bmatrix} 1 \\ 2 \\ \vdots \\ N \end{bmatrix}$

$$\text{Distance} = \frac{|c_1 - c_2|}{\|\bar{a}\|}$$

$$= \frac{|1-(-1)|}{\sqrt{1^2 + 2^2 + \dots + N^2}} = \frac{2}{\sqrt{\frac{N(N+1)(2N+1)}{6}}}$$

$$= \frac{2\sqrt{6}}{\sqrt{N(N+1)(2N+1)}}$$

Problem 2 : Consider the two hyperplanes

$$C_0 : \bar{a}^T \bar{x} + b = 1$$

$$C_1 : \bar{a}^T \bar{x} + b = -1$$

What is the distance between them ?

$$\Rightarrow \bar{a}^T \bar{x} = 1 - b$$

$$\bar{a}^T \bar{x} = -1 - b$$

$c_1 = 1 - b$, $c_2 = -1 - b$

$$\text{Distance} = \frac{|c_1 - c_2|}{\|\bar{a}\|} = \frac{|(1-b)-(-1-b)|}{\|\bar{a}\|} = \frac{2}{\|\bar{a}\|}$$

Maximum Margin classifier

The problem to determine classifier with the maximum margin is

$$\max \frac{2}{\|\bar{a}\|_2} = \min \|\bar{a}\|_2$$

$$\bar{a}^T \bar{x}_i + b \geq 1 , \quad \bar{x}_i \in C_0$$

$$\bar{a}^T \bar{x}_i + b \leq -1 , \quad \bar{x}_i \in C_1$$

The above problem is <u>convex</u> and can be readily solved. This classifier is termed as <u>Support vector Classifier</u> (SVC). Also termed as <u>Support vector Machine</u> (SVM)

<u>Problem 3</u> : For the given data below, formulate the SVM problem.

$$\bar{x}_1 = \begin{bmatrix} -1 \\ -2 \end{bmatrix}, \quad \bar{x}_2 = \begin{bmatrix} -3 \\ -1 \end{bmatrix} \quad \in C_0$$

$$\bar{x}_3 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \quad \bar{x}_4 = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \quad \in C_1$$

The SVM problem is $\min \|\bar{a}\|$

$$\Rightarrow \begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} -1 \\ -2 \end{bmatrix} + b \geq 1 \quad \Rightarrow -a_1 - 2a_2 + b \geq 1$$

$$\begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} -3 \\ -1 \end{bmatrix} + b \geq 1 \quad \Rightarrow -3a_1 - a_2 + b \geq 1$$

$$\begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix} + b \leq -1 \quad \Rightarrow 2a_1 + 2a_2 + b \leq -1$$

$$\begin{bmatrix} a_1 & a_2 \end{bmatrix} \begin{bmatrix} 3 \\ 4 \end{bmatrix} + b \leq -1 \quad \Rightarrow 3a_1 + 4a_2 + b \leq -1$$

<u>Dual SVM</u>

Let λ_i denote <u>Lagrange Multipliers</u>.

Dual SVM problem can be formulated as

$$\max \sum_{i=1}^{2M} \lambda_i - \frac{1}{2} \sum_{i,j=1}^{2M} \lambda_i \lambda_j y_i y_j \bar{x}_i^T \bar{x}_j$$

subject to $\lambda_i \geq 0$

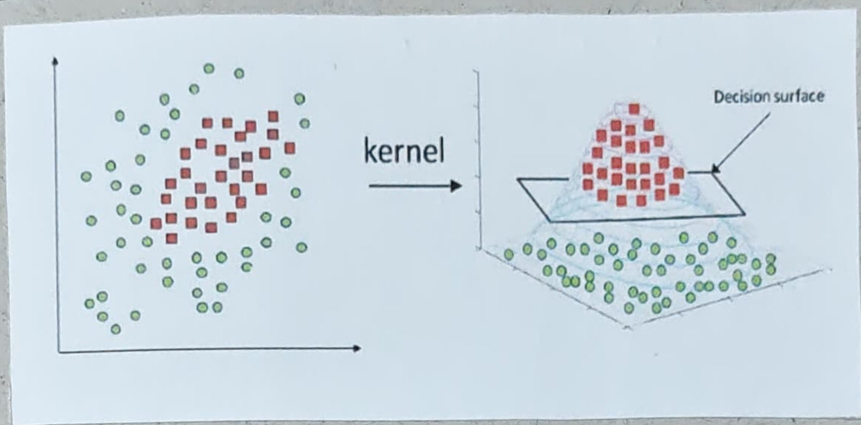
$$\sum_{i=1}^{2M} \lambda_i y_i = 0$$

Kernel SVM

One can now replace $\bar{x}_i^T \bar{x}_j$ by a kernel $K(\bar{x}_i, \bar{x}_j)$. This can be used to model non-linear features. The most popular is the Gaussian kernel, defined as

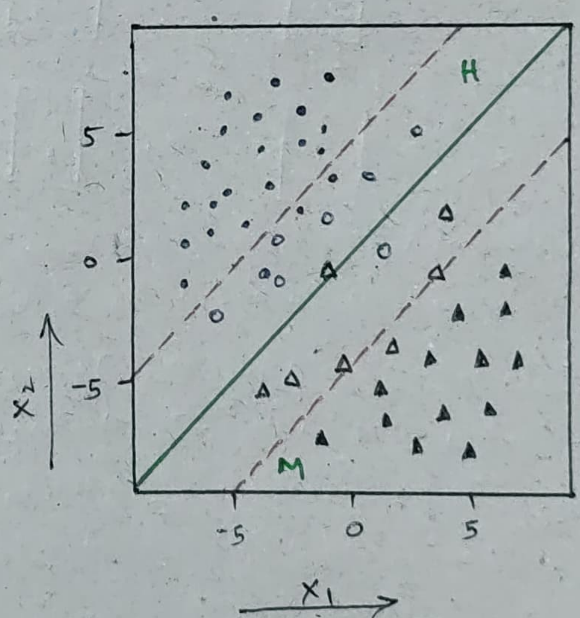
$$K(\bar{x}_i, \bar{x}_j) = \exp\left(- \frac{\|\bar{x}_i - \bar{x}_j\|^2}{2\sigma^2}\right)$$

Non-linear kernel is suited where linear separation is not possible.



Approximate Classifier

When the points are not linearly separable, one can employ an approximate classifier. This leads to classification error.



The Approximate classifier minimizes the classification error.

Mathematically, the approximate classifier can be represented as

$$\bar{a}^T \bar{x}_i + b \geq 1 - u_i \quad , \quad \bar{x}_i \in C_0$$

$$\bar{a}^T \bar{x}_i + b \leq -1 + v_i \quad , \quad \bar{x}_i \in C_1$$

where,

$u_i \geq 0$, $v_i \geq 0$ are <u>slack variables</u>.

Hence, the '<u>SOFT' classifier</u> problem is given as

$$\min \sum_{i=1}^{N} u_i + \sum_{i=1}^{N} v_i$$

$$\bar{a}^T \bar{x}_i + b \geq 1 - u_i \quad , \quad \bar{x}_i \in C_0$$

$$\bar{a}^T \bar{x}_i + b \leq -1 + v_i \quad , \quad \bar{x}_i \in C_1$$

$$u_i \geq 0 \ , \ v_i \geq 0$$

<u>Problem 4</u>: Consider the data below and formulate the soft classifier problem.

$$\bar{x}_1 = \begin{bmatrix} -1 \\ -2 \end{bmatrix}, \ \bar{x}_2 = \begin{bmatrix} -3 \\ -1 \end{bmatrix} \in C_0$$

$$\bar{x}_3 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \ \bar{x}_4 = \begin{bmatrix} 3 \\ 4 \end{bmatrix} \in C_1$$

The SOFT classifier problem is

$$\min \quad u_1 + u_2 + v_1 + v_2$$

$$\Rightarrow \quad -a_1 - 2a_2 + b \geq 1 - u_1$$

$$-3a_1 - a_2 + b \geq 1 - u_2$$

$$2a_1 + 2a_2 + b \leq -1 + v_1$$

$$3a_1 + 4a_2 + b \leq -1 + v_2$$

$$u_1 \geq 0 \ ; \ u_2 \geq 0 \ , \ v_1 \geq 0 \ , \ v_2 \geq 0$$

Classifier design

The classifier can be trained as

$$\bar{a}^T \bar{x}_i + b \geq 0 \;, \quad \bar{x}_i \in C_0$$

$$\bar{a}^T \bar{x}_i + b \leq 0 \;, \quad \bar{x}_i \in C_1$$

Need to determine $\bar{a}$ and b that characterize the linear classifier.

$$\bar{a}^T \bar{x}_i + b \geq 0 \;, \quad \bar{x}_i \in C_0$$

$$\bar{a}^T \bar{x}_1 + b \leq 0 \;, \quad \bar{x}_i \in C_1$$

The above problem has a trivial solution !

$$\bar{a} = 0 \quad \text{and} \quad b = 0$$

Therefore, problem has to be modified.