eMasters in Communication Systems

Prof. Aditya Jagannatham

Elective Module:

Advanced ML Techniques

Chapter 4

Support Vector Machines

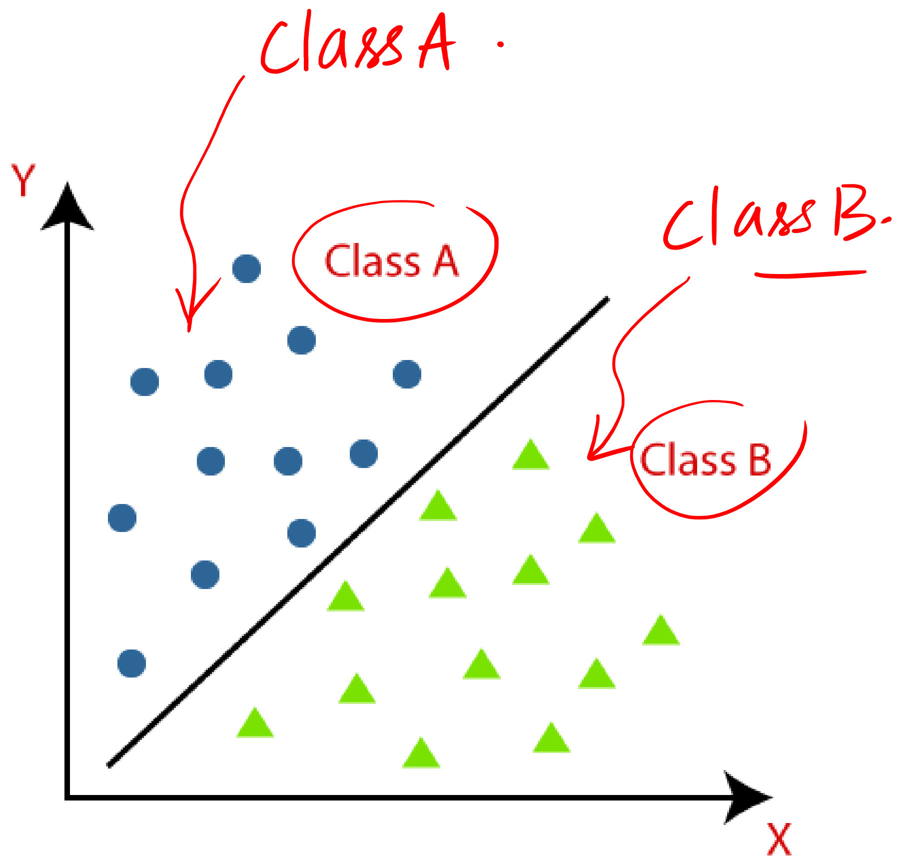
SVM.

Classification

- **Classification** is an important tool in ML.
  - *Determines to which* ***class*** *an observation belongs*

Binary Classification

- **Binary Classification**
  - ⇒ 2 CLASSES.



Binary Classification

- ***Binary Classification***
  - $\Rightarrow$2 Classes

Binary Classification
- *Binary Classification*
  - $\Rightarrow$2 Classes

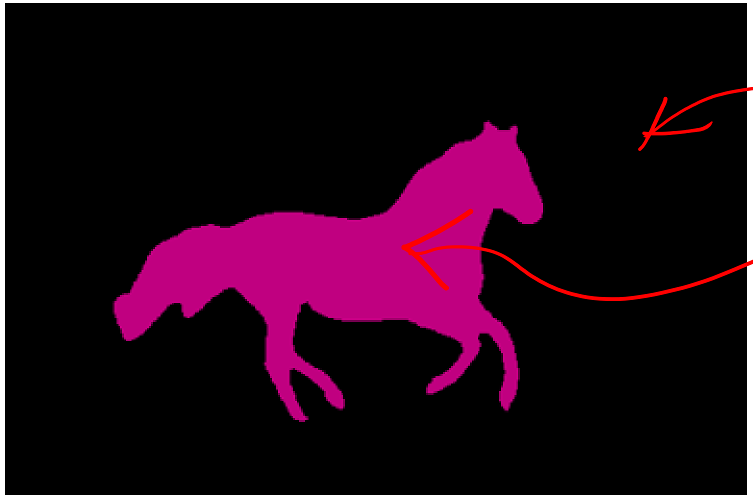
Applications

- **Image segmentation**
  - Classify pixels as belonging to

Background & Foreground.

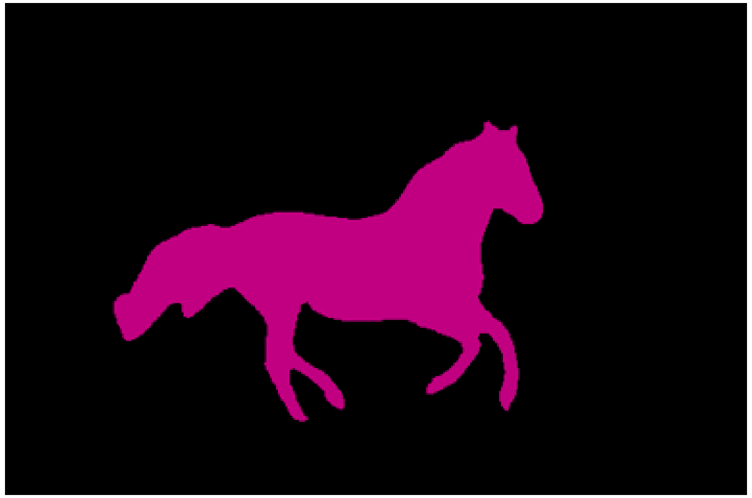
BACKGROUND.

FOREGROUND.



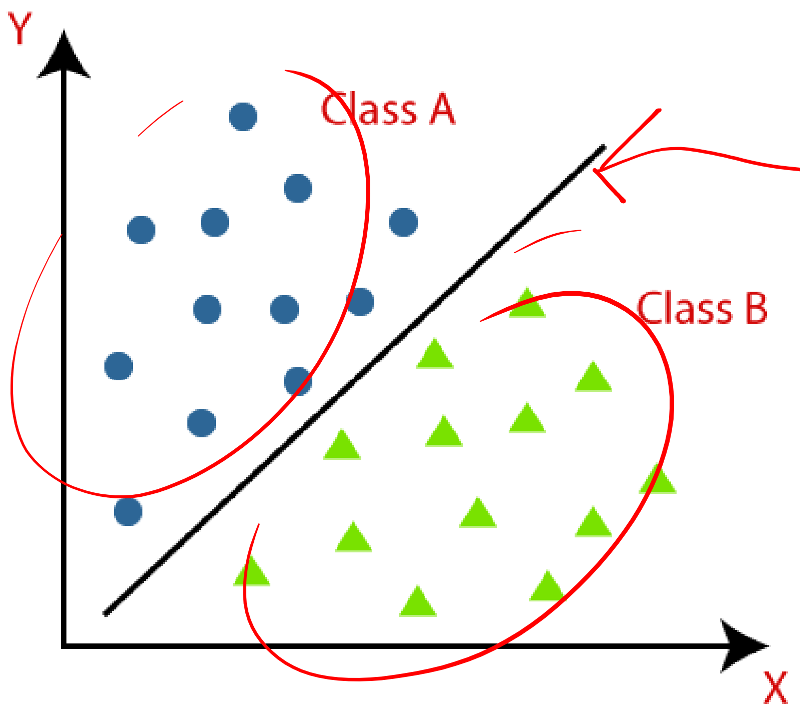
Applications

- **Image segmentation**
  - Classify pixels as belonging to **foreground** or **background**



Linear classifier

- Linear classifier corresponds to a hyperplane in N dimensions
  - Easy to determine and analyse!



2D Straight Line

3D Plane

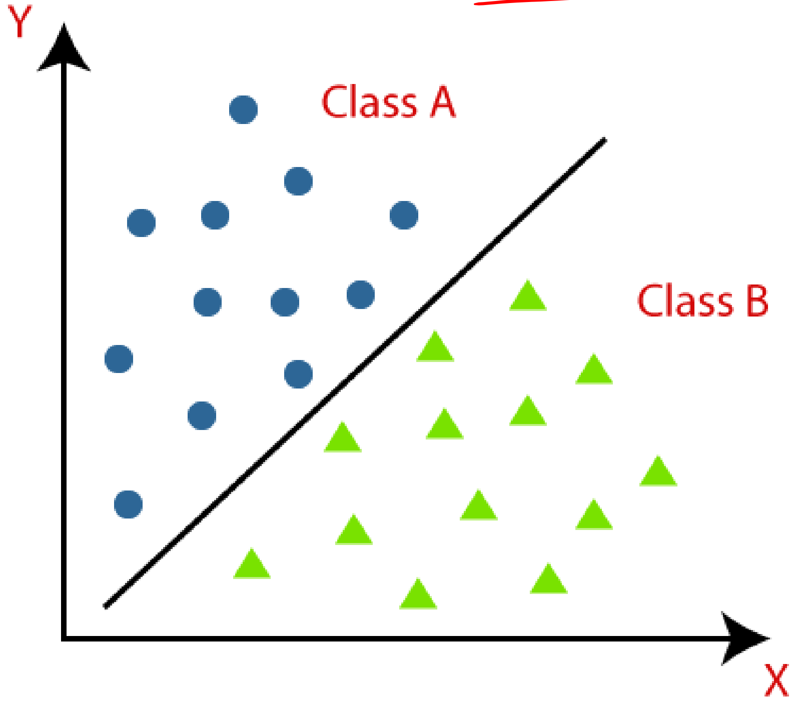
N.D N dimensions

hyperplane

Linear classifier

Compute ·
Learn

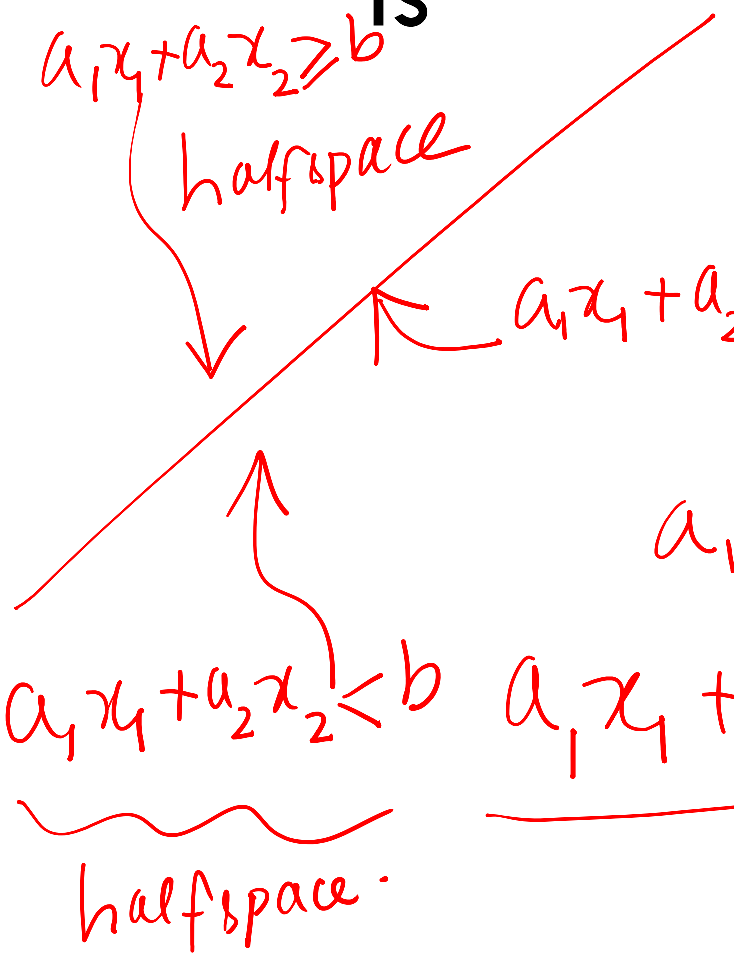
- Linear classifier corresponds to a **hyperplane** in N dimensions
  - Easy to determine and analyse!



Linear classifier

$$\underbrace{a_1x_1 + a_2x_2 + \cdots + a_Nx_N = b}_{\text{hyperplane}}$$

- General structure of a linear classifier is



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

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

halfspace

$a_1x_1 + a_2a_2 + \cdots + a_Nx_N \geq b$

$$\underbrace{a_1x_1 + a_2x_2 + \cdots + a_Nx_N < b}_{\text{halfspace.}}$$

Linear classifier

- General structure of a linear classifier is

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

$$C_1 : \bar{\mathbf{a}}^T \bar{\mathbf{x}} < b$$

Foreground.

C_0: halfspace.

C_1: Background.

halfspace.

Disease detection

$\bar{x}$ → C_0: disease absent

$\bar{x}$ → C_1: disease present.

Linear classifier

Linear classifier

- How to determine the linear classifier
- Consider the **training set**

$$\left.\begin{array}{l} \overbrace{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_M}^{M\text{ points.}} \in C_0 \\ \underbrace{\bar{x}_{M+1}, \bar{x}_{M+2}, \ldots, \bar{x}_{2M}}_{M\text{ points.}} \in C_1 \end{array}\right\} \text{Training data.}$$

Linear classifier

M points. C_0

- How to determine the linear classifier
- Consider the **training set**

$$\underbrace{\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \dots, \bar{\mathbf{x}}_M}_{} \in C_0$$

C_1

$$\underbrace{\bar{\mathbf{x}}_{M+1}, \bar{\mathbf{x}}_{M+2}, \dots, \bar{\mathbf{x}}_{2M}}_{\text{M points.}} \in C_1$$

Classifier design

- The **classifier** can be trained as

$$\bar{a}^T \bar{x}_i + b \geq 0 \qquad C_0 \quad i = 1, 2, \dots, M$$

$$\bar{a}^T \bar{x}_i + b < 0 \qquad C_1 \quad i = M+1, \dots, 2M$$

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

$$\bar{a}^T \bar{x} + b = 0$$

Classifier design

- The **classifier** can be trained as

$$\bar{\mathbf{a}}^T \bar{\mathbf{x}}_i + b \geq 0, 1 \leq i \leq M$$

$$\bar{\mathbf{a}}^T \bar{\mathbf{x}}_i + b \leq 0, M + 1 \leq i \leq 2M$$

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

Classifier design

$$\bar{\mathbf{a}}^T \bar{\mathbf{x}}_i + b \geq 0, 1 \leq i \leq M$$

$$\bar{\mathbf{a}}^T \bar{\mathbf{x}}_i + b \leq 0, M + 1 \leq i \leq 2M$$

- What is the problem with this **formulation**?

Classifier design

$$\bar{\mathbf{a}}^T \bar{\mathbf{x}}_i + b \geq 0, 1 \leq i \leq M$$

$$\bar{\mathbf{a}}^T \bar{\mathbf{x}}_i + b \leq 0, M + 1 \leq i \leq 2M$$

- The above problem has a **trivial solution**!

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

$\bar{a} = 0, b = 0$

Trivial Solution

Classifier design

- Therefore, problem has to be ***modified***

To avoid trivial solution

Modified optimization problem C_0

- The **modified optimization problem** is

2 Parallel hyperplanes:

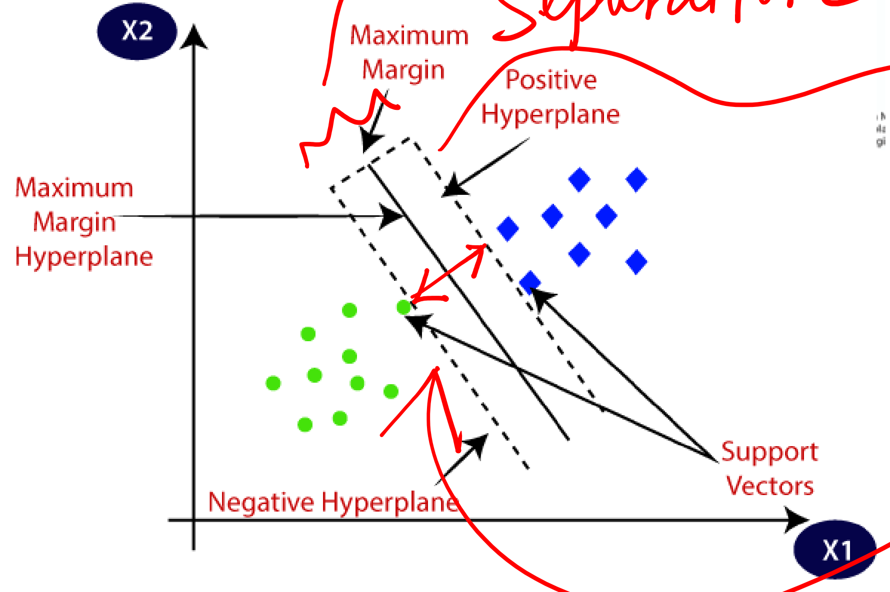
$$\begin{cases} \bar{a}^T \bar{x}_i + b \geq 1 & i = 1, 2, \ldots, M \\ \bar{a}^T \bar{x}_i + b \leq -1 & i = M+1, \ldots, 2M \end{cases}$$

C_1

Margin Separation

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

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



Modified optimization problem

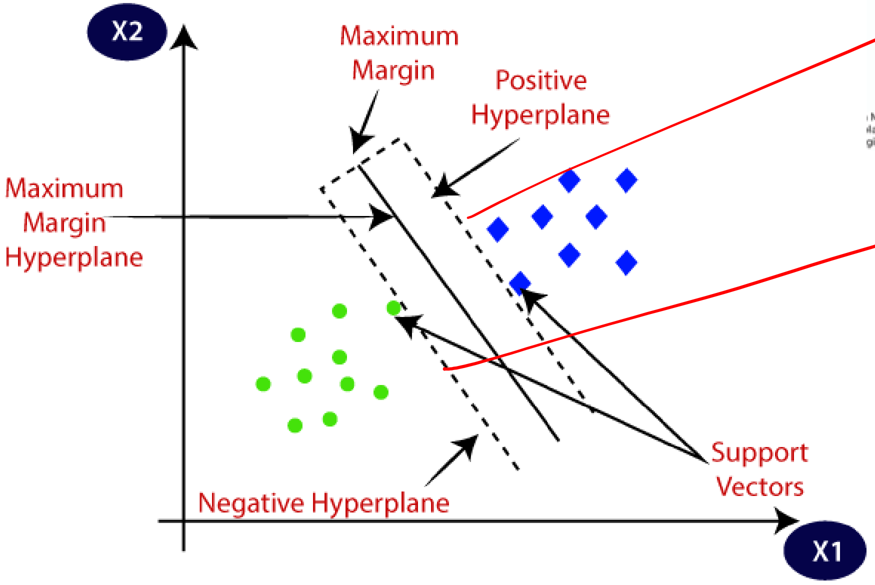
- The **modified optimization problem** is

$$\bar{\mathbf{a}}^T \bar{\mathbf{x}}_i + b \geq 1,\ 1 \leq i \leq M$$

$$\bar{\mathbf{a}}^T \bar{\mathbf{x}}_i + b \leq -1,\ M+1 \leq i \leq 2M$$

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

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

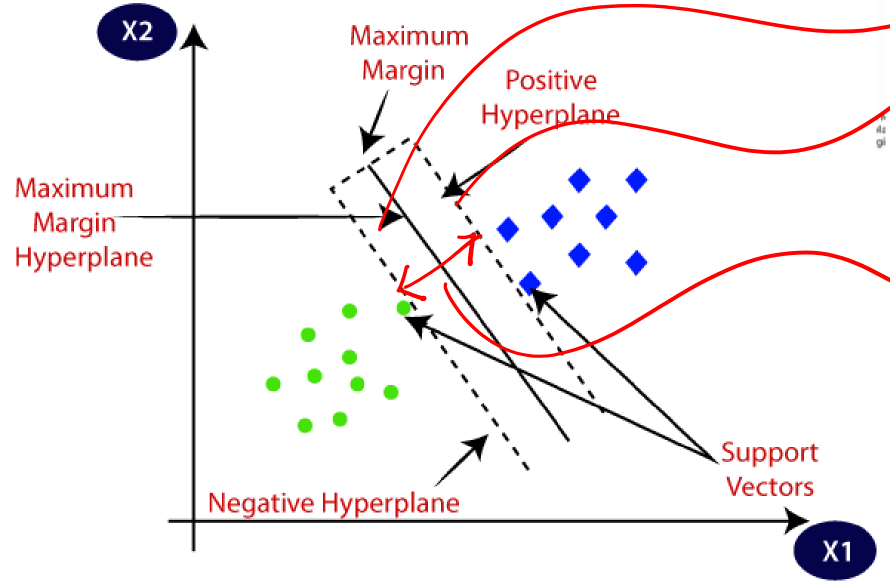


Modified optimization problem

- Determines two Parallel hyperplanes.
- Separates both classes by a Slab.

$\bar{a}^T \bar{x} + b = c$

For different values of c are parallel.

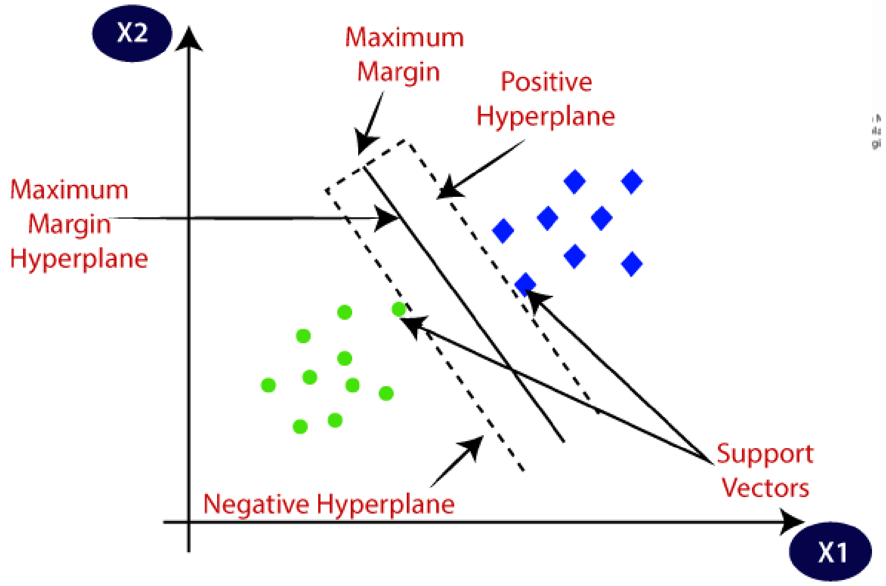


parallel hyperplanes.

Margin Slab.

Modified optimization problem

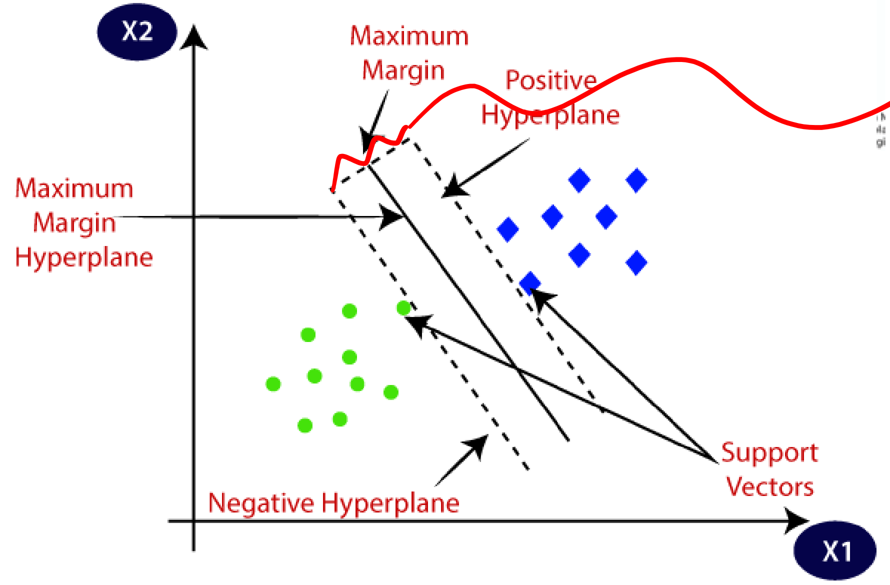
- Determines two ***parallel hyperplanes***
- Separates both classes by a **slab**



Modified optimization problem

- The width of the slab is termed the MARGIN
- Best classifier maximizes the margin between the two classes

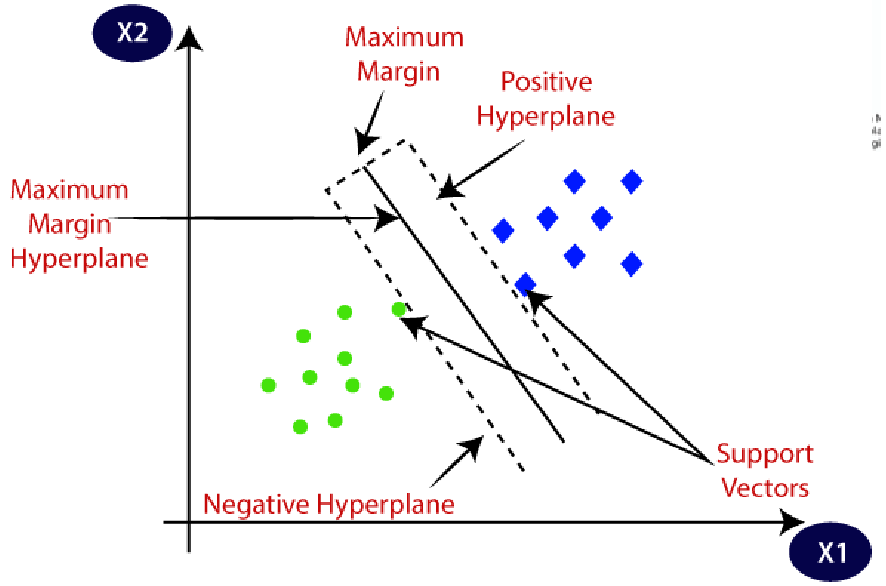
misclassification error is small.



Margin

Modified optimization problem

- The width of the slab is termed the **margin**
- Best classifier **maximizes the margin** between the two classes



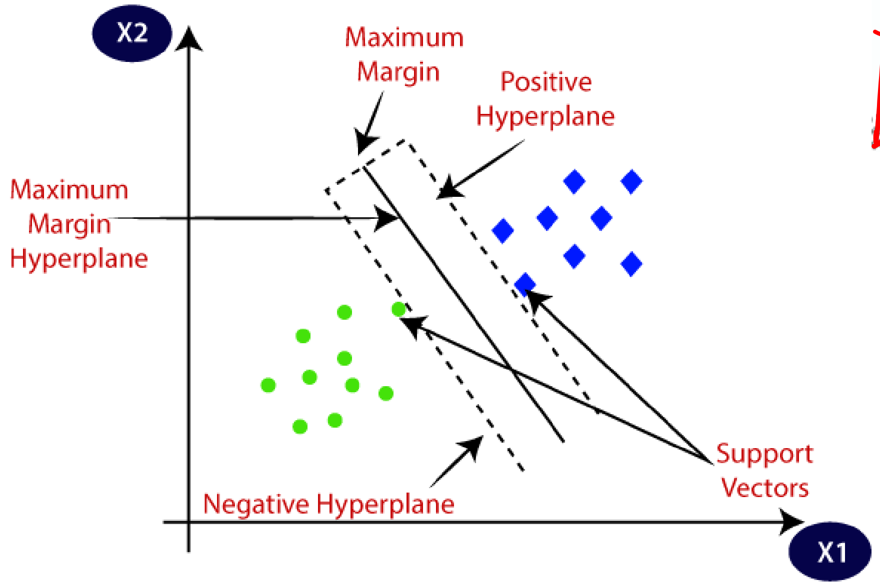
Margin

$$\|\bar{a}\| = \sqrt{a_1^2 + a_2^2 + \cdots + a_N^2}$$

- How to determine the **margin** between two hyperplanes?

$$\bar{a}^T\bar{x} = c_1$$
$$\bar{a}^T\bar{x} = c_2$$

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

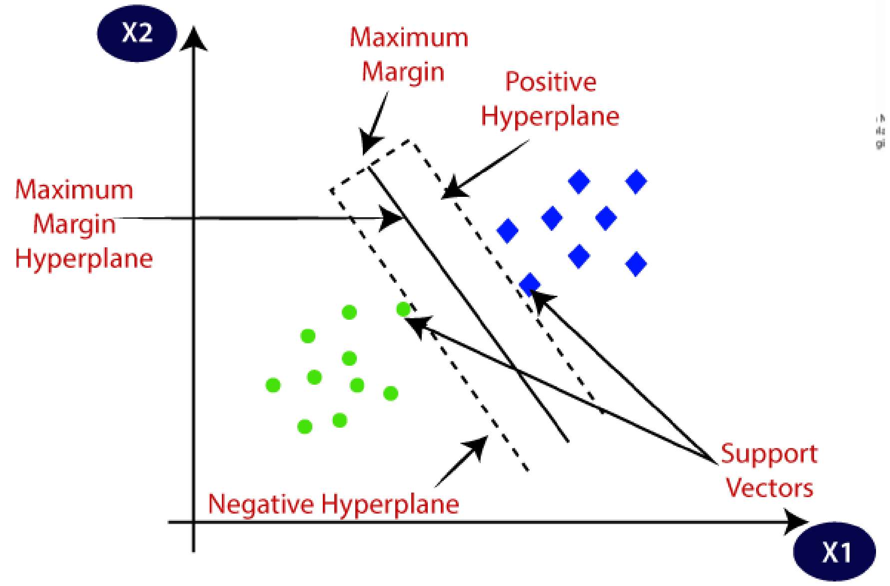


Margin

- How to determine the **margin** between two hyperplanes?

Parallel hyperplanes-

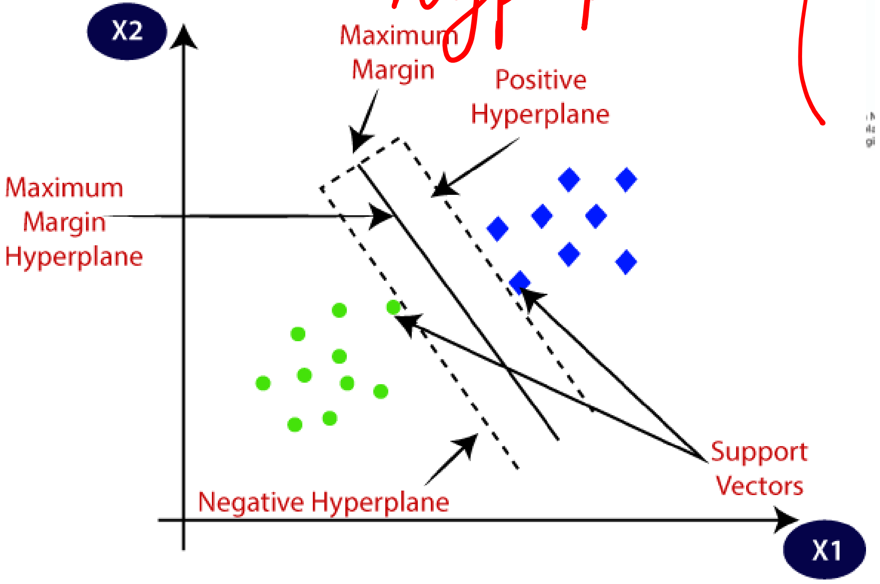
$$\bar{\mathbf{a}}^T\bar{\mathbf{x}} = c_1$$
$$\bar{\mathbf{a}}^T\bar{\mathbf{x}} = c_2$$



Margin

- Distance between the two hyperplanes is

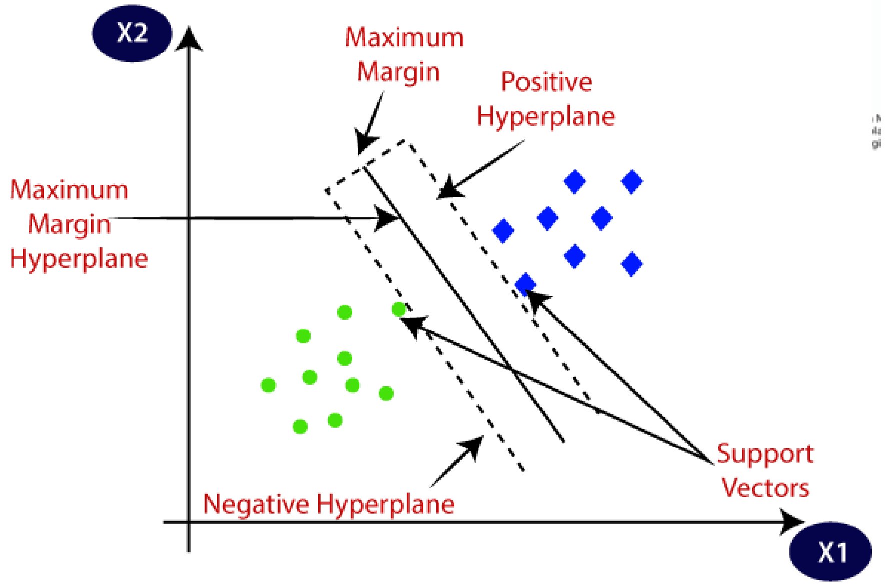
Margin distance between hyperplanes $\left\{ \dfrac{|c_1 - c_2|}{\|\bar{a}\|} \right.$



Margin

- Distance between the two hyperplanes is

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



Example

- What is the distance between the **hyperplanes**

$$x_1 + 2x_2 + 3x_3 + \cdots + Nx_N = 1 \quad c_1$$

$$x_1 + 2x_2 + 3x_3 + \cdots + Nx_N = -1 \quad c_2$$

$$\bar{a} = \begin{bmatrix} 1 \\ 2 \\ \vdots \\ N \end{bmatrix}$$

$$d = \frac{2}{\sqrt{1^2 + 2^2 + \cdots + N^2}}$$

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

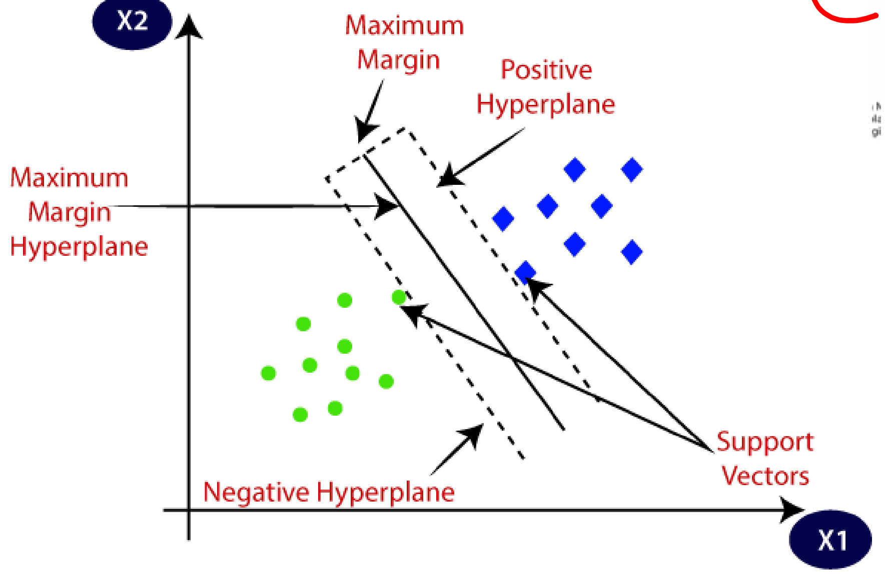
Margin

- Consider the two ***hyperplanes***

$$\ell_0 : \bar{a}^T\bar{x}+b=1 \Rightarrow \bar{a}^T\bar{x}=1-b \quad c_1.$$

$$\ell_1 : \bar{a}^T\bar{x}+b=-1 \Rightarrow \bar{a}^T\bar{x}=-1-b \quad c_2$$

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

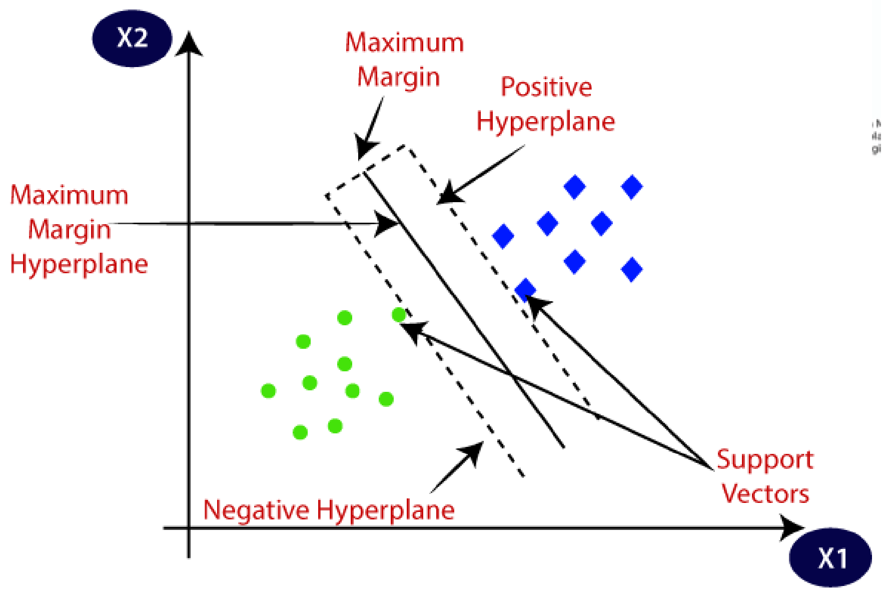


Margin

- Consider the two *hyperplanes*

$$C_0 : \bar{\mathbf{a}}^T \bar{\mathbf{x}} = 1 - \mathrm{b}$$

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

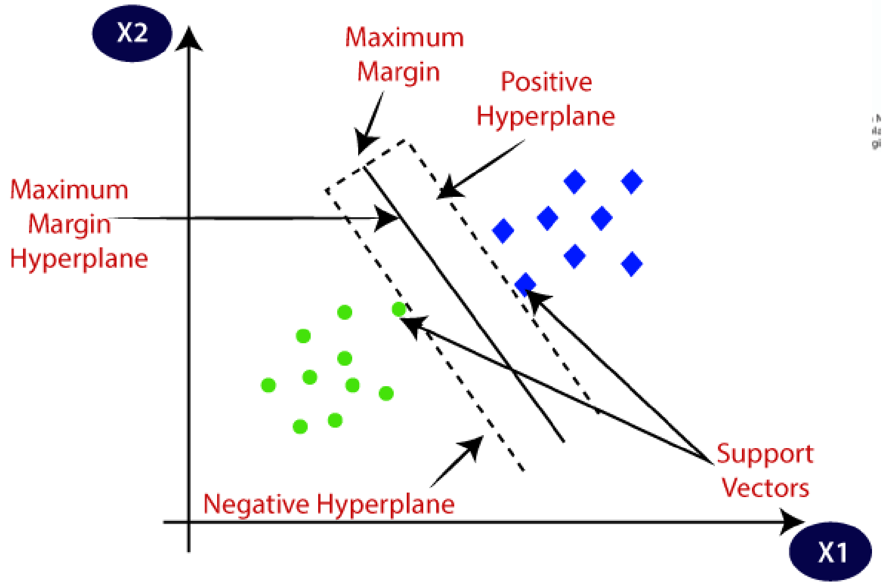


Margin

- Distance between the hyperplanes or the **margin** is

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

$$= \frac{2}{\|\bar{a}\|}.$$



Margin

- Distance between the hyperplanes or the **margin** is

$$\frac{2}{\|\bar{a}\|}$$

In Hard classifier NO classification error

NO points in margin or slab region.

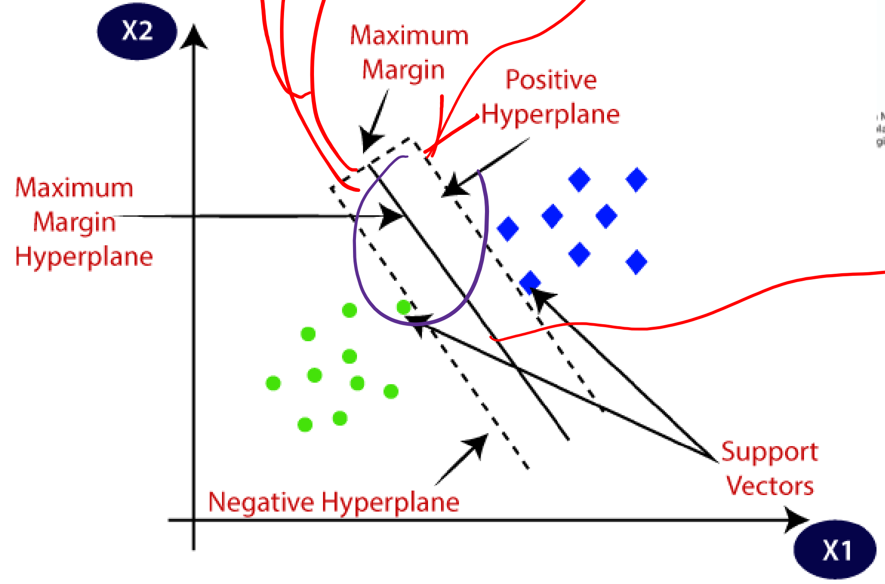
Margin between positive & negative hyperplanes.

maximize margin

Optimal classifier

Hard classifier

$\min \|\bar{a}\|$



Maximum margin classifier

- The problem to determine classifier with **maximum margin** is

convex

$$\sqrt{a_1^2 + a_2^2 + \cdots + a_N^2}$$

$$\min \ \|\bar{a}\|$$

$$\text{s.t.} \begin{cases} \bar{a}^T \bar{x}_i + b \geq 1 & i = 1, \ldots, M \\ \bar{a}^T \bar{x}_i + b \leq -1 & i = M+1, \ldots, 2M \end{cases}$$

Very Easily

Affine constraints.

$\Rightarrow$ convex optimization problem.

Maximum margin classifier

- The problem to determine classifier with **maximum margin** is

Objective function

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

2M Training Set

constraints

$$C_0: \bar{\mathbf{a}}^T \bar{\mathbf{x}}_i + b \geq 1, 1 \leq i \leq M$$

$$C_1: \bar{\mathbf{a}}^T \bar{\mathbf{x}}_i + b \leq -1, M + 1 \leq i \leq 2M$$

CONVEX OPTIMIZATION PROBLEM.

Support vector machine

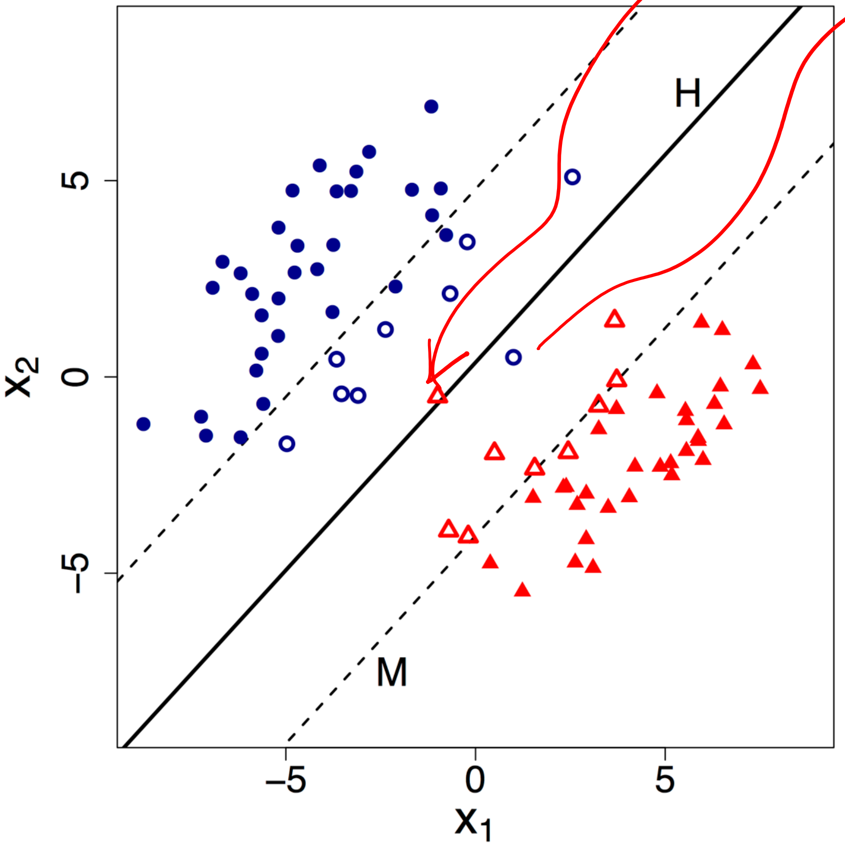
- The above problem is **convex** and can be readily solved
- This classifier is termed a **Support Vector Machine (SVM)**

Efficiently

Approximate classifier

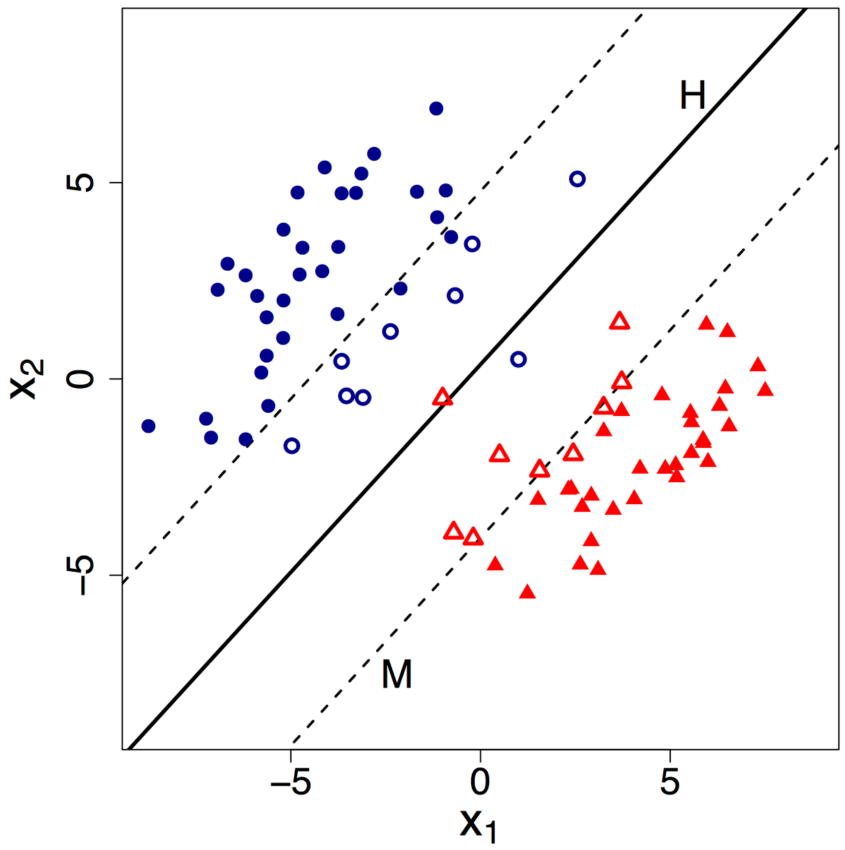
- When the points are not linearly separable.

Classification error



Approximate classifier

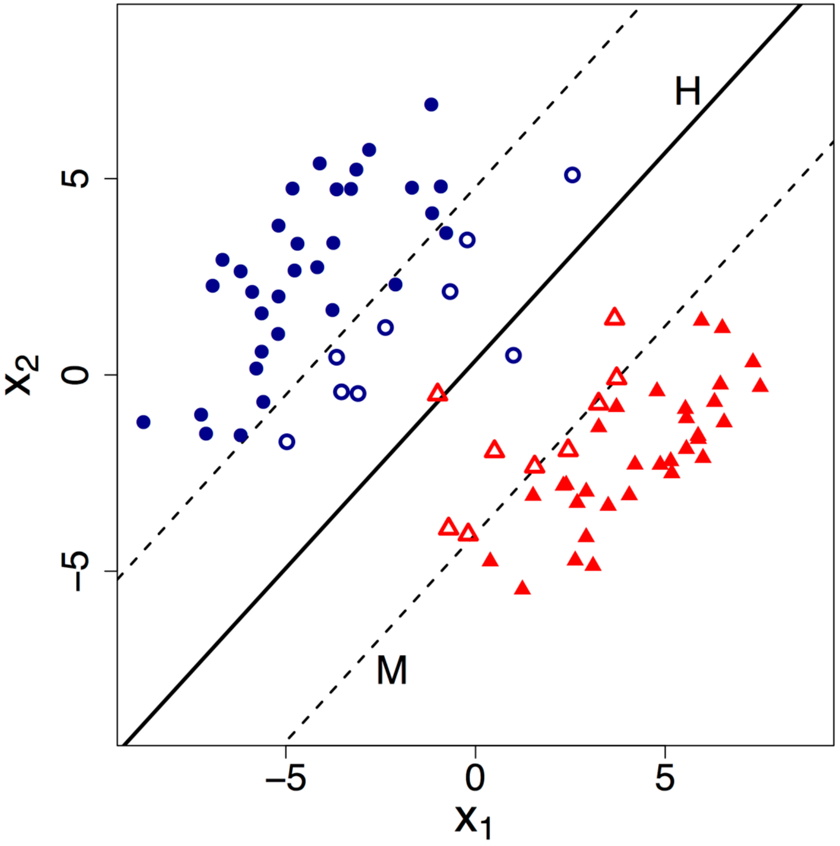
- When the points are not **linearly separable**



Approximate classifier

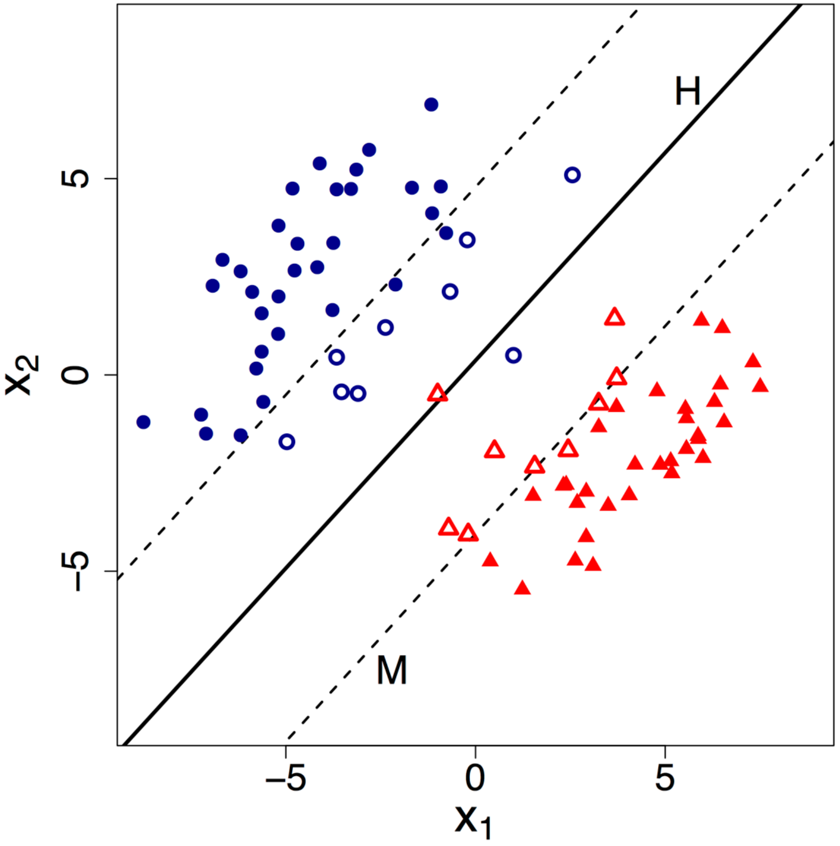
- One can employ an **approximate classifier**

Tolerate some classification error



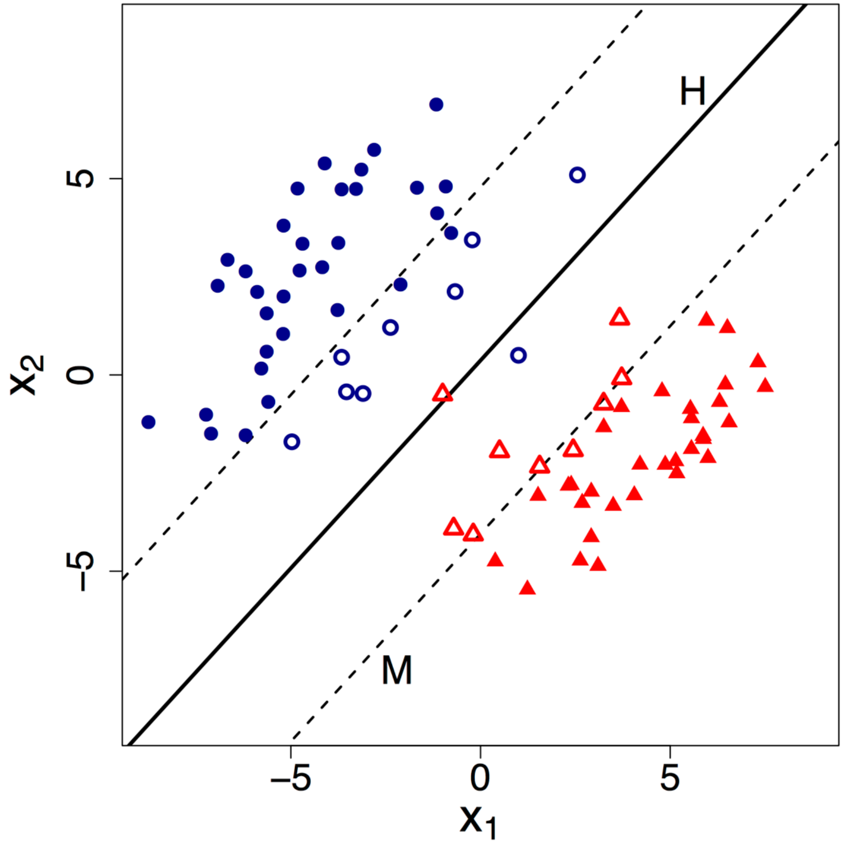
Approximate classifier

- Since the points are not **linearly separable** one needs to tolerate some error



Approximate classifier

- The Approximate Classifier **minimizes** the Classification Error



Approximate classifier

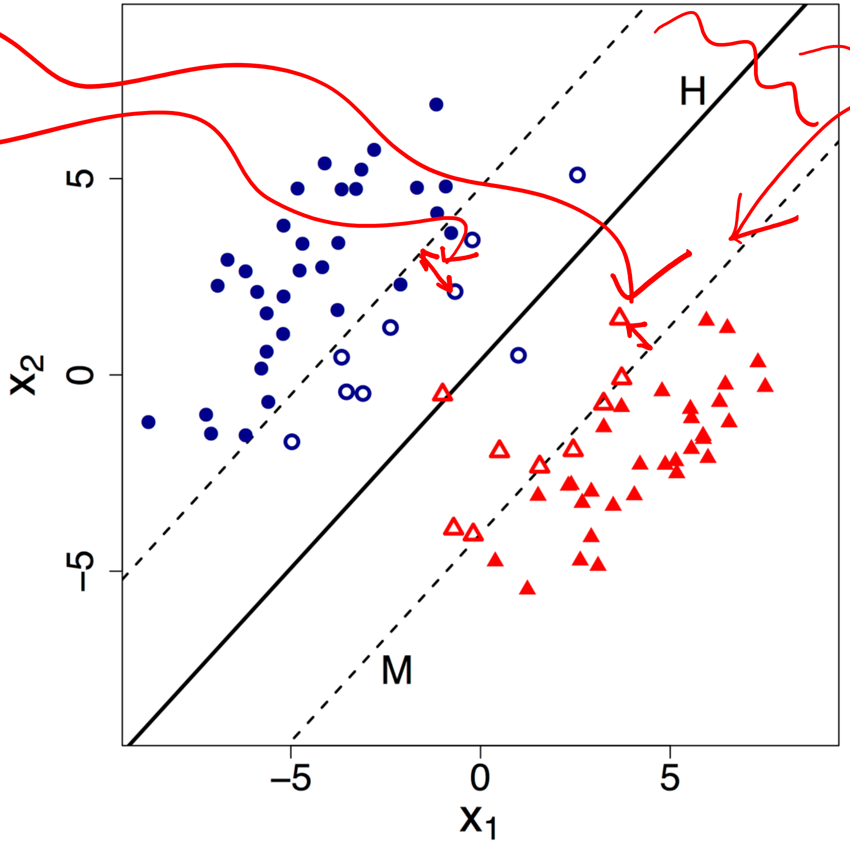
- The Approximate Classifier **minimizes** the **classification error**

'Slack'

+ve hyperplane

-ve hyperplane

No man's land.



Approximate classifier

$u_i \geq 0$ }
$v_i \geq 0$ } Slack has to be non-negative

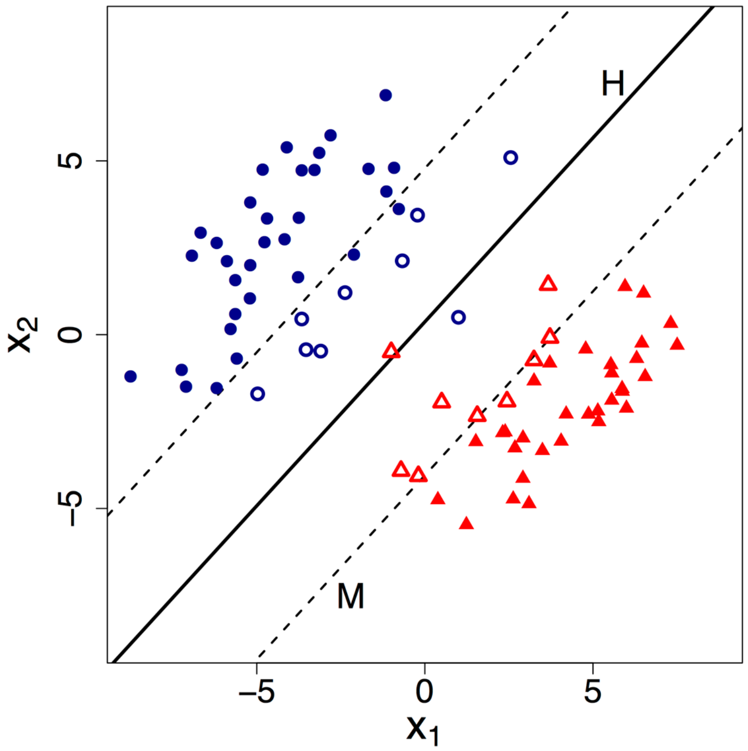
- Mathematically this can be represented as

$$\bar{a}^T \bar{x}_i + b \geq 1 - u_i \quad i = 1, 2, \ldots, M$$

$$\bar{a}^T \bar{x}_i + b \leq -1 + v_i \quad i = M+1, \ldots, 2M$$

u_i, v_i : Slack

$u_i, v_i \geq 0$



Approximate classifier

$\mathcal{C}_0$

- Mathematically this can be represented as

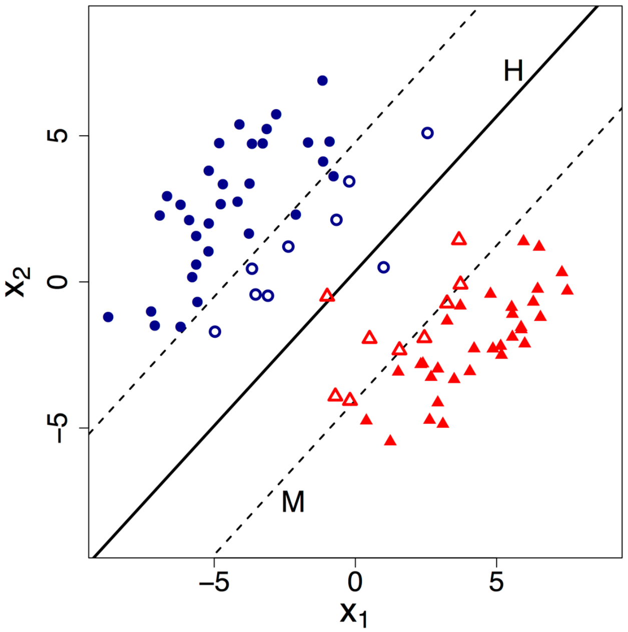
Slacks $\mathcal{C}_0$.

$$\bar{\mathbf{a}}^T\bar{\mathbf{x}}_i + b \geq 1 - u_i, 1 \leq i \leq M$$

$$\bar{\mathbf{a}}^T\bar{\mathbf{x}}_i + b \leq -1 + v_i, M + 1 \leq i \leq 2M$$

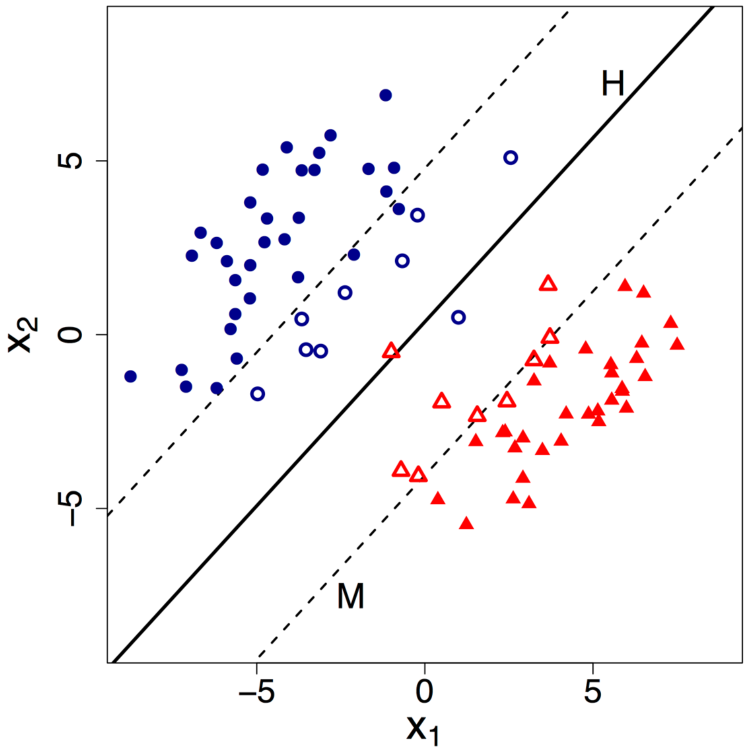
Slacks $\mathcal{C}_1$

$\mathcal{C}_1$



Approximate classifier

- $u_i \geq 0$, $v_i \geq 0$, are Slack variables.



- Relaxation
- Deviation

Approximate classifier

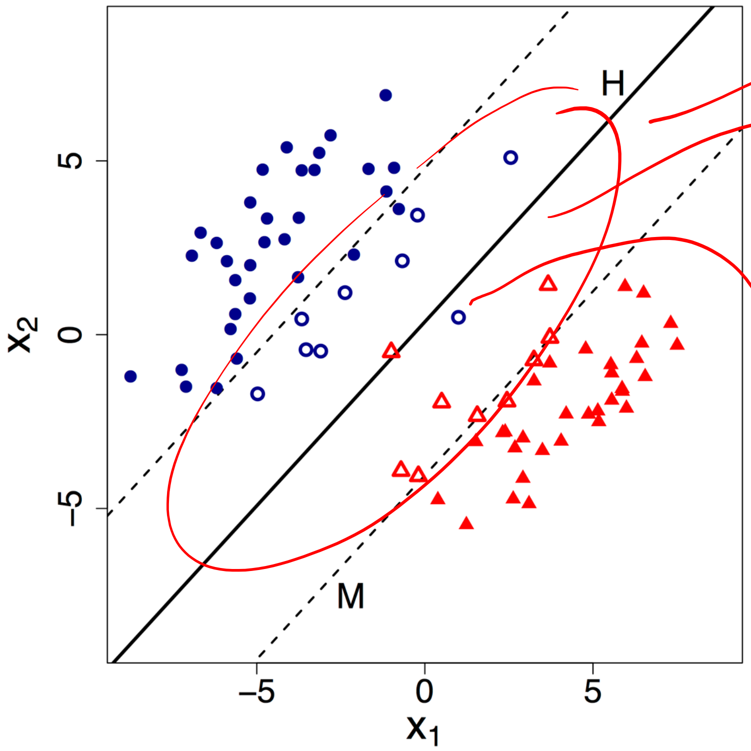
- $u_i \geq 0$, $v_i \geq 0$, are ***slack variables***

Extent to which they penetrate slab.

TOlerate some classification error

SOFT classifier

There are some points in margin.



Approximate classifier

Linear program.

- Hence, the **'SOFT' classifier** problem is given as

minimize Total Slack

Objective
Total slack.
Linear

$$\sum_i u_i + \sum_i v_i$$

Non-negativity of slack

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

Linear classification constraints.

Constraints.
Affine

$$\bar{a}^T \bar{x}_i + b \geq 1 - u_i \quad i = 1, \ldots, M$$

$$\bar{a}^T \bar{x}_i + b \leq -1 + v_i, \quad i = M+1, \ldots, 2M$$

Convex optimization problem.

Approximate classifier

- Hence, the **'SOFT' classifier** problem is given as

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

$$\bar{\mathbf{a}}^T \bar{\mathbf{x}}_i + b \geq 1 - u_i, 1 \leq i \leq M$$

$$\bar{\mathbf{a}}^T \bar{\mathbf{x}}_i + b \leq -1 + v_i, M + 1 \leq i \leq 2M$$

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

SOFT Classifier

Instructors may use this white area (14.5 cm / 25.4 cm) for the text.
Three options provided below for the font size.

Font: Avenir (Book), Size: 32, Colour: Dark Grey

Font: Avenir (Book), Size: 28, Colour: Dark Grey

Font: Avenir (Book), Size: 24, Colour: Dark Grey

Do not use the space below.