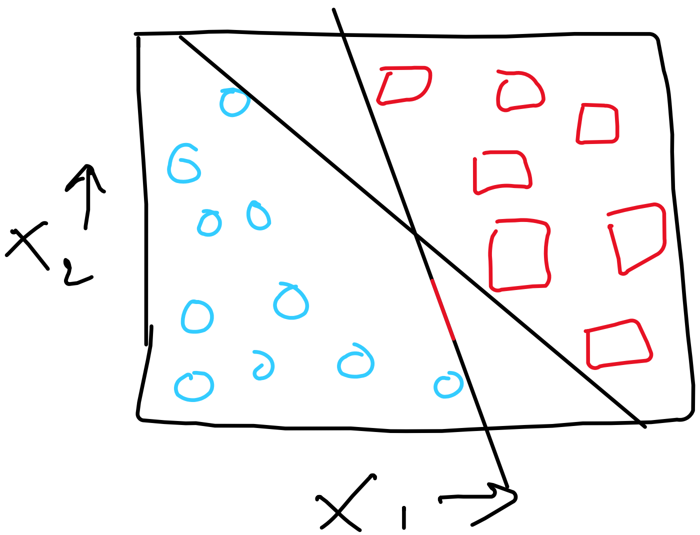
Support vector machine

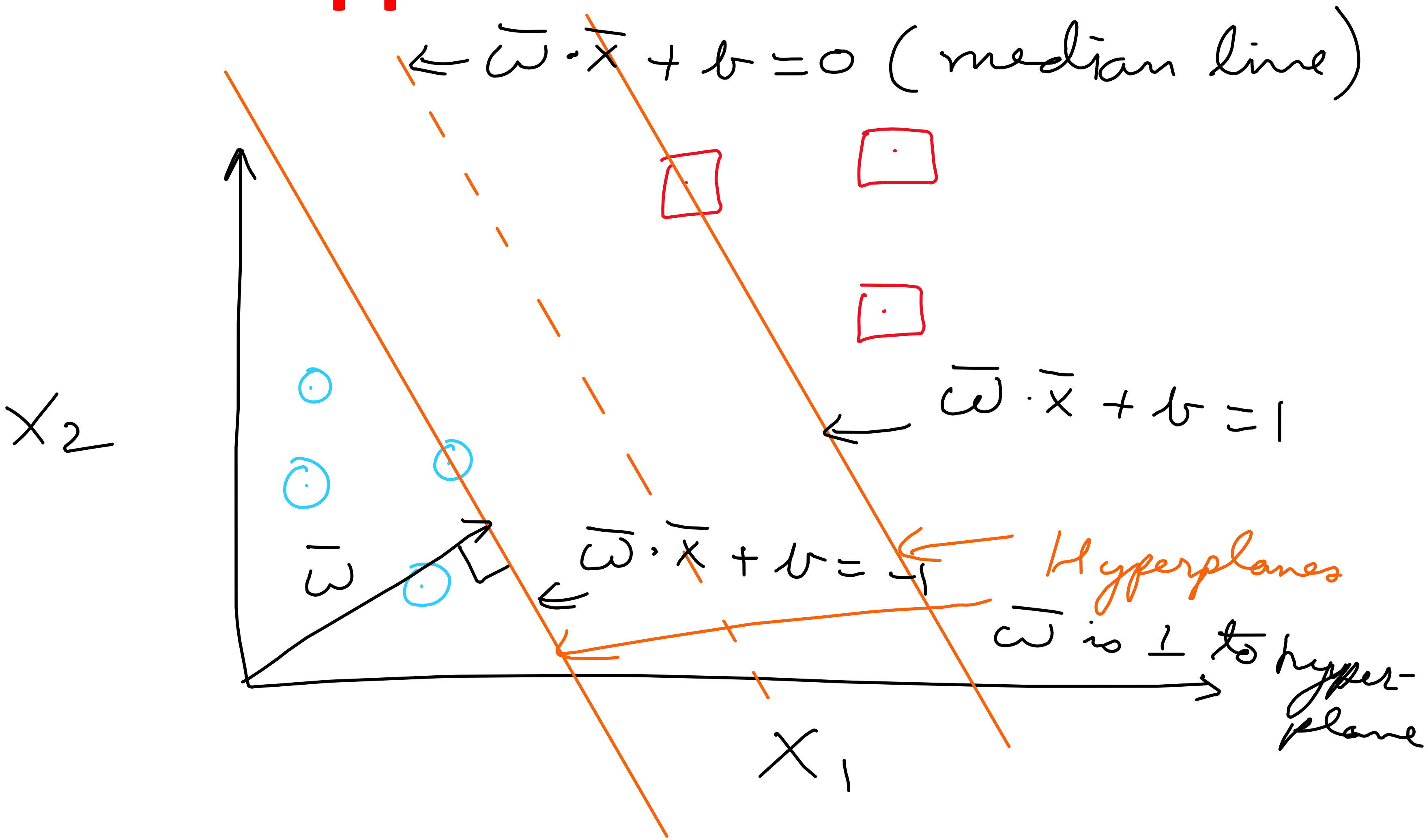
Prof. Asim Tewari
IIT Bombay

Support Vector Classifier

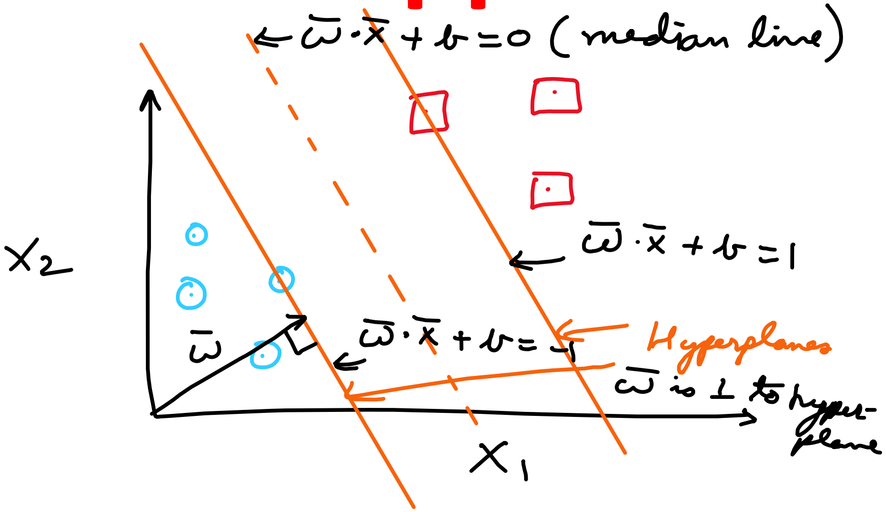


SVC : Maximum - margin hyperplane

Support Vector Classifier



Support Vector Classifier



1. Any point above or on $\bar{w}\cdot\bar{x}+b=1$ belongs to □

2. Any point on or below $\bar{w}\cdot\bar{x}+b=-1$ belongs to ○

For "□": $\bar{w}\cdot\bar{x}+b \geq 1$

For "○" $\bar{w}\cdot\bar{x}+b \leq -1$

For an unknown $\bar{u}$ the decision rule is based on maximum-margin hyperplane.

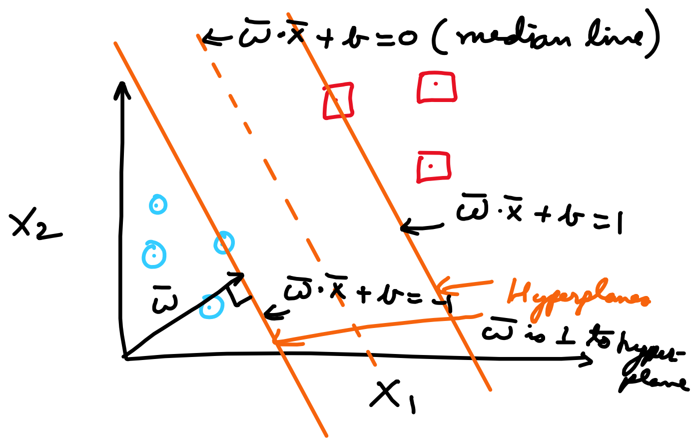
$\bar{w}\cdot\bar{u}+b \geq 0 \Rightarrow$ □

$\bar{w}\cdot\bar{u}+b \leq 0 \Rightarrow$ ○

Support Vector Classifier

Standardize data set

For point i



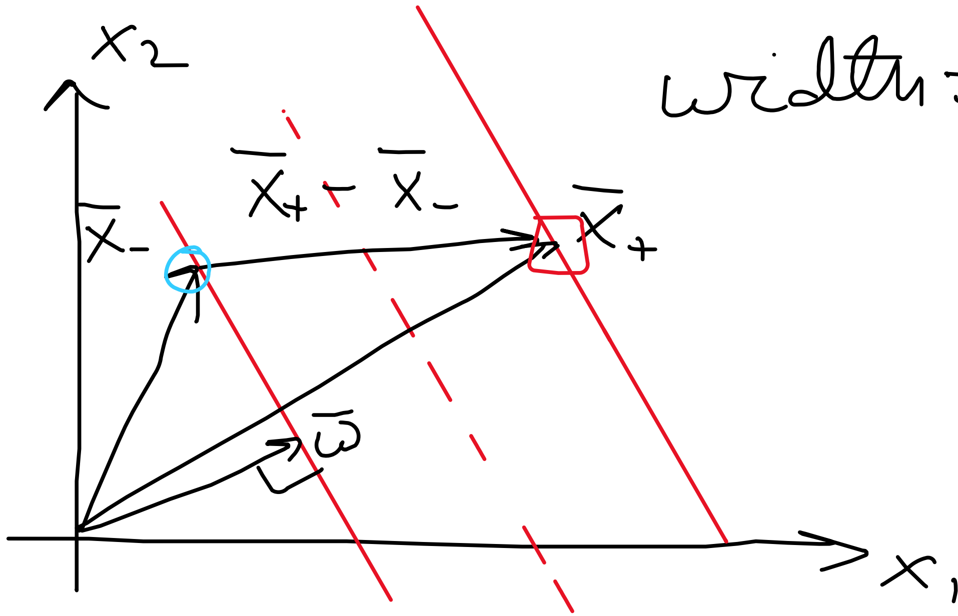
$$y_i = \begin{cases} 1 & \text{if it belongs to } \square \\ 0 & \text{if it belongs to } \bigcirc \end{cases}$$

$$y_i(\bar{w}\cdot\bar{x}_i + b) \geq 1 \quad \begin{cases} \square & \text{True for} \\ \bigcirc & \text{both} \end{cases}$$

$$\Rightarrow y_i(\bar{w}\cdot\bar{x}_i + b) - 1 \geq 0$$

$$y_i(\bar{w}_i\cdot\bar{x}_i + b) - 1 = 0$$

Support Vector Classifier



$$\text{width} = (\bar{x}_+ - \bar{x}_-) \cdot \frac{\bar{\omega}}{\|\bar{\omega}\|}$$

↑

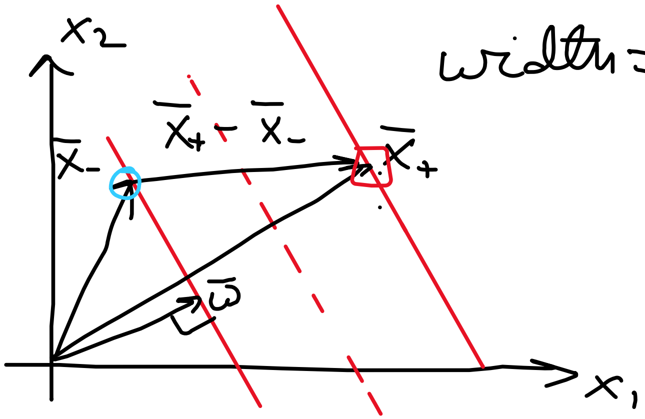
unit vector

$$y_i(\bar{\omega}\cdot\bar{x}_i + b) - 1 = 0$$

we get $\bar{\omega}\cdot\bar{x}_+ = 1 - b$ and $\bar{\omega}\cdot\bar{x}_- = -1 - b$

$$\text{width} = (\bar{x}_+ - \bar{x}_-) \cdot \frac{\bar{\omega}}{\|\bar{\omega}\|} = \frac{\bar{x}_+\cdot\bar{\omega} - \bar{x}_-\cdot\bar{\omega}}{\|\bar{\omega}\|}$$

Support Vector Classifier



$$\text{width} = \frac{\bar{x}_+ \cdot \bar{w} - \bar{x}_- \cdot \bar{w}}{\|\bar{w}\|}$$

$$= \frac{(1 - b) - (-1 - b)}{\|\bar{w}\|}$$

$$\Rightarrow \text{width} = \frac{2}{\|\bar{w}\|}$$

So we want two hyperplanes $y_i(\bar{w} \cdot \bar{x}_i + b) - 1 = 0$ such that the width $\left(\frac{2}{\|w\|}\right)$ is maximum.

Maximize $\dfrac{2}{\|w\|}$

$\Rightarrow$ Minimize $\|w\|/2$

Support Vector Classifier

$\equiv$ minimize $\|\omega\|^2$

Minimize $\|\omega\|/2$ with a constrain

$$y_i(\bar{\omega}\cdot\bar{x}_i + b) - 1 = 0$$

$$L = \frac{1}{2}\|\bar{\omega}\|^2 - \sum \alpha_i\left[y_i(\bar{\omega}\cdot\bar{x}_i + b) - 1\right]$$

$$\min_{\omega, b}(L) \Rightarrow \frac{\partial L}{\partial \bar{\omega}} = 0, \quad \frac{\partial L}{\partial b} = 0$$

$$\Rightarrow \frac{\partial L}{\partial \omega} = \frac{\partial}{\partial \bar{\omega}}\left[\frac{1}{2}\|\omega\|^2\right] - \frac{\partial}{\partial \bar{\omega}}\left[\sum \alpha_i y_i(\omega\cdot x_i + b) - 1\right]$$

$$\frac{\partial L}{\partial \omega} = 0 \Rightarrow \bar{\omega} = \sum \alpha_i y_i \bar{x}_i$$

Support Vector Classifier

$$L = \frac{1}{2}\|\bar{\omega}\|^2 - \sum \alpha_i \left[y_i (\bar{\omega} \cdot x_i + b) - 1 \right]$$

$$\frac{\partial L}{\partial \bar{\omega}} = 0 \Rightarrow \bar{\omega} = \sum (\alpha_i y_i \bar{x}_i)$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow -\sum \alpha_i y_i = 0 \text{ or } \sum \alpha_i y_i = 0$$

$$\therefore L = \frac{1}{2} \underbrace{\Big(\sum \alpha_i y_i \bar{x}_i\Big) \cdot \Big(\sum \alpha_j y_j \bar{x}_j\Big)}_{\bar{\omega} \cdot \bar{\omega} = \|\bar{\omega}\|^2}$$

$$- \sum \Big[\alpha_i \Big(y_i \big(\underbrace{\sum \alpha_j y_j x_j}_{\bar{\omega}} \big) \cdot \bar{x}_i + b \Big) - 1 \Big]$$

Support Vector Classifier

$$L = \frac{1}{2}\left(\sum \alpha_i y_i \bar{x}_i\right)\cdot\left(\sum \alpha_j y_j \bar{x}_j\right)$$

$$- \left(\sum \alpha_i y_i \bar{x}_i\right)\cdot\left(\sum \alpha_j y_j \bar{x}_j\right)$$

$$- \sum \alpha_i y_i b + \sum \alpha_i$$

$$\Rightarrow L = \sum \alpha_i - \frac{1}{2}\sum_i \sum_j \left(\alpha_i \alpha_j y_i y_j \bar{x}_i \cdot \bar{x}_j\right)$$

$$L = f\left(\bar{x}_i \cdot \bar{x}_j\right)$$

dot product

Support Vector Classifier

Now if $\sum \alpha_i y_i \underbrace{\bar{x}_i \cdot \bar{u}} + b \geq 0$

Decision Rule

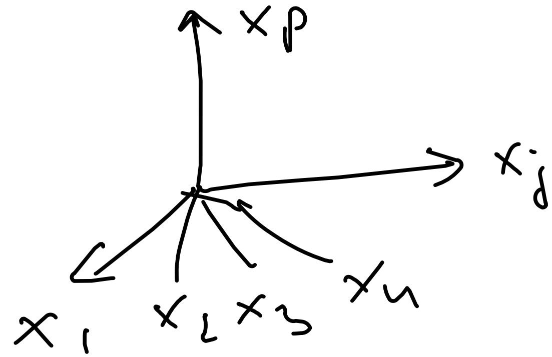
$+$

$-$

$\bar{x}_i \cdot \bar{u}$

also depends on this "." dot product.

Dimensionality reduction



If p is very large then how can you reduce the calculation of distance measure

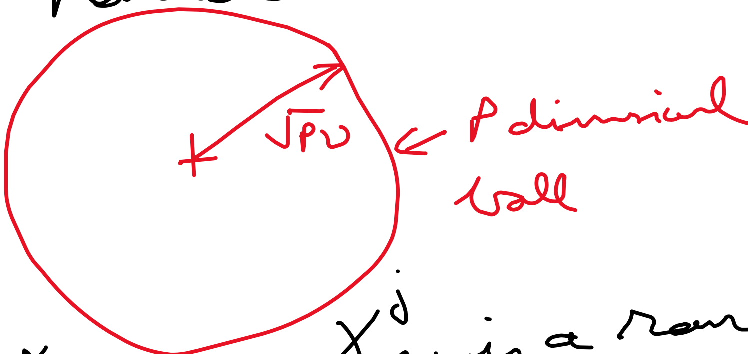
Eg: $\bar{X}_1 \xleftrightarrow{\text{distance}} \bar{X}_2$

$p = 10^6$

Dimensionality reduction

Random vectors: $\bar{X}_{R_1} = (x_{R_1}^1, x_{R_1}^2, \dots \dots x_{R_1}^p)$

$\bar{X}_{R_2} = (x_{R_2}^1, x_{R_2}^2, \dots \dots x_{R_2}^p)$



* $X_{R_i}^j$ is a random number with mean zero and variance of ν.

* All $x_{R_i}^j$ are i.i.d

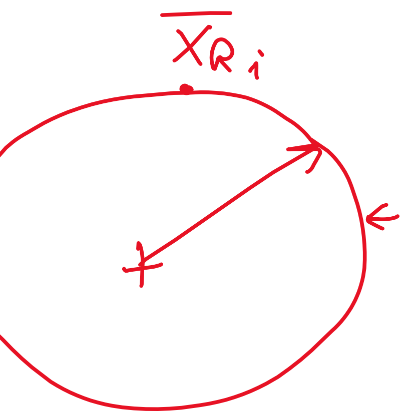
If p is very large then $\|\bar{X}_{R_i}\| \approx \sqrt{p\nu}$

$$\|\bar{X}_{R_i}\| = \sqrt{\underbrace{(x_{R_i}^1)^2}_{\nu} + \underbrace{(x_{R_i}^2)^2}_{\nu} + \dots \underbrace{(x_{R_i}^p)^2}_{\nu}} = \sqrt{p\nu}$$

Dimensionality reduction

Random vectors: $\overline{X}_{R_1} = (x^1_{R_1}, x^2_{R_1}, \ldots \ldots x^p_{R_1})$

$$\overline{X}_{R_2} = (x^1_{R_2}, x^2_{R_2} \ldots \ldots x^p_{R_2})$$



Thin shell on the p-dimensional Ball

$\overline{X}_{R_i}$ are iid $N(0, \nu)$

$$\overline{X}_{R_1} \cdot \overline{X}_{R_2} = \left(x^1_{R_1} \cdot x^1_{R_2} + x^2_{R_1} \cdot x^2_{R_2} + \ldots \ldots x^p_{R_1} \cdot x^p_{R_2}\right)$$

$+ \quad - \quad + \quad + \ldots$

For very large p

$\overline{X}_{R_1} \cdot \overline{X}_{R_2} \simeq 0 \Rightarrow \overline{X}_{R_1} \perp \overline{X}_{R_2}$

Data is of p dimension ($\sim 10^6$) then create N (say 10^3) random vectors and use them as basis to cast data from p to N dimensions.