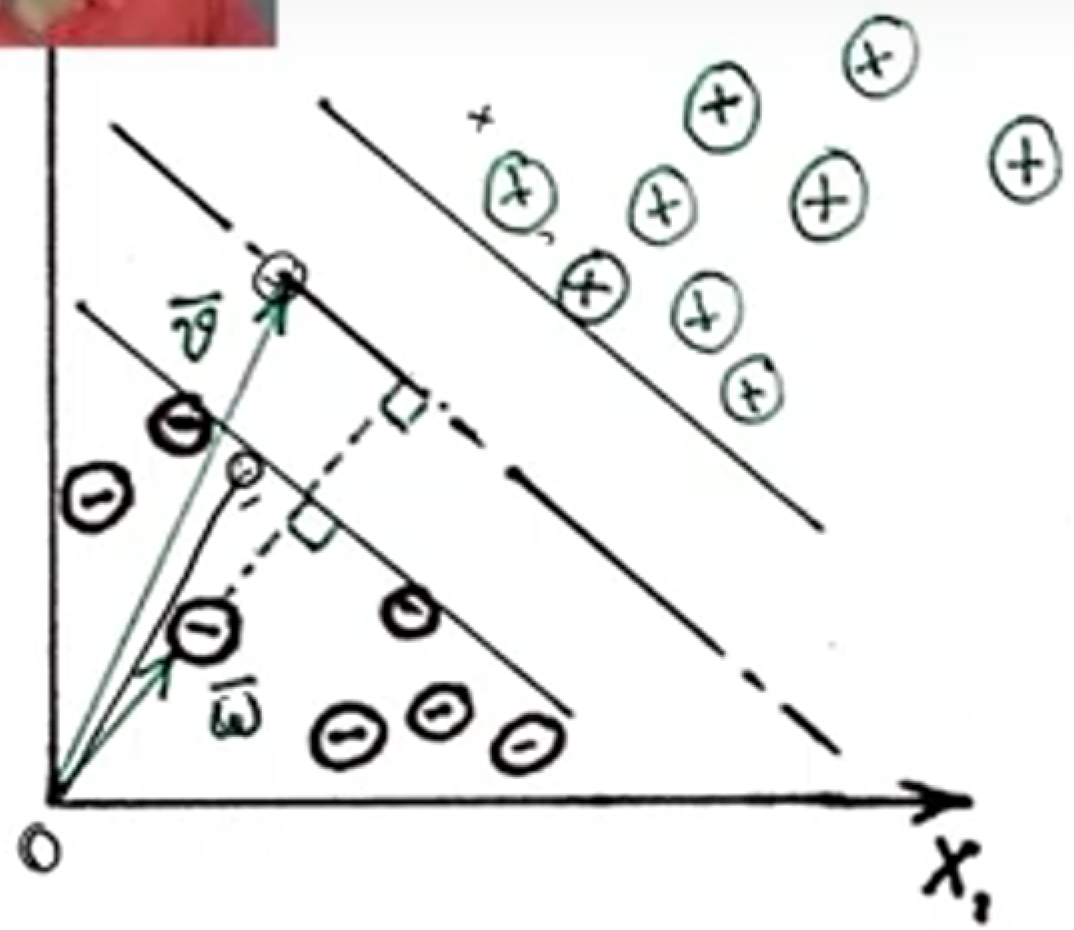
# Support Vector Machine(SVM)

Vapnik

- It is a binary Linear Classifier which classifies data with maximum possible geometric margin
- Credit of inventing SVM primarily goes to Vladimir Vapnik (1974,1995), although many other researchers /developer worked on it to give the Algorithm practical shape usable in the industry.
- We will be discussing the following:
  - SVM as a maximum margin classifier.
  - Formulating classification problem as constrained quadratic optimization problem
  - Primal Problem Solution
  - Understanding the problem by solving it by hand
  - Dual problem formulation and why do we require it?
  - KKT condition
  - Concept of Kernel function
  - SMO algorithm ( A SVM training Algorithm)

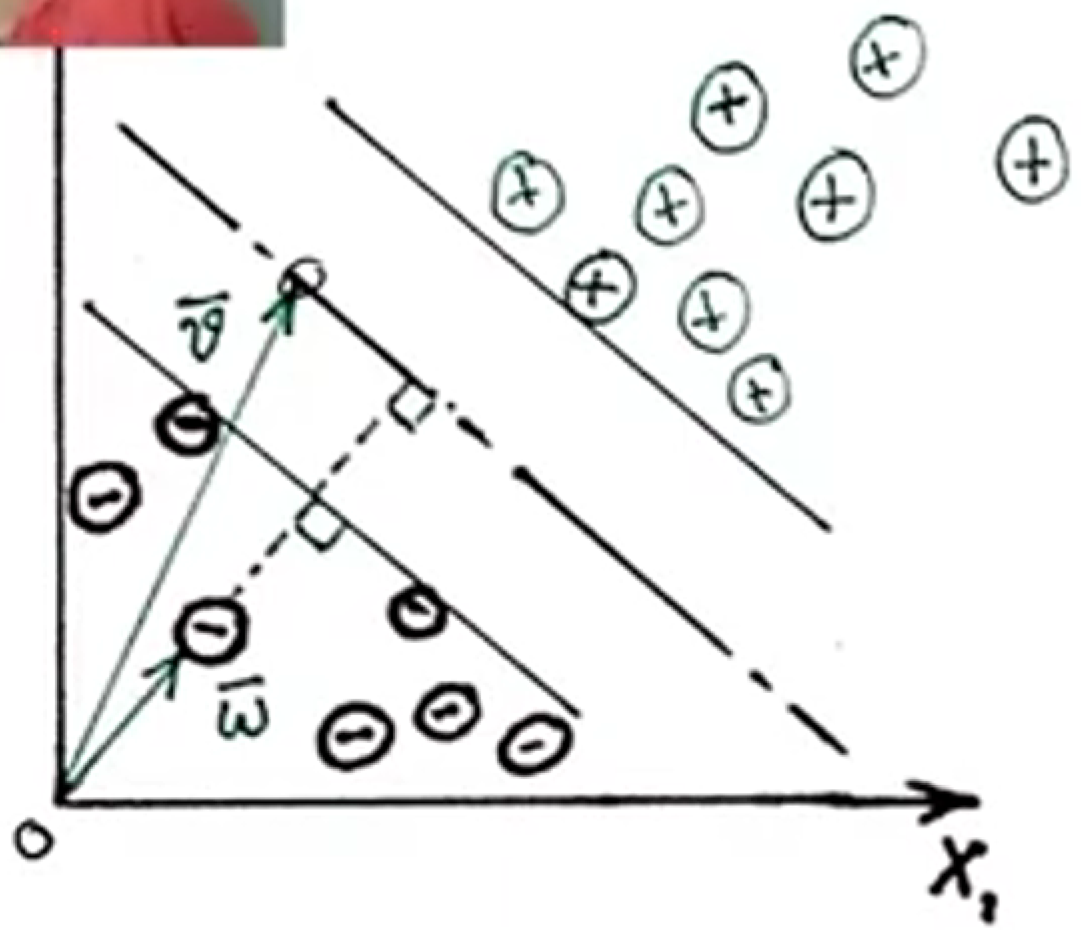
Introduction to Machine Learning- Lecture 30 ( Introducing Support Vector Machine)




$\vec{v} \cdot \vec{w} \geq b$, then the arbitrary sample belongs to + ve class

otherwise it belongs to − ve class.

$< b$




1. Trying to make a decision rule:

$$Y^{(i)} \in \{\underset{0}{-1}, \underset{1}{1}\}$$

$\bar{v}\cdot\bar{w} \geq K$, k any Constant $= -b$

$\bar{v}\cdot\bar{w} + b \geq 0$ then +ve class

$\bar{v}\cdot\bar{w} + b < 0$ then -ve class

} Decision rule (1)

2. Try to develop contraints:-

Take a +ve sample $\bar{s}_+$

$$\bar{w}\cdot\bar{s}_+ + b \geq 1$$

$\bar{w}\cdot\bar{s}_+ + b$ must be greater than 1 ———— (2)

Take a -ve sample $\bar{s}_-$ ———— (3)

$$\bar{w}\cdot\bar{s}_- + b \leq -1$$

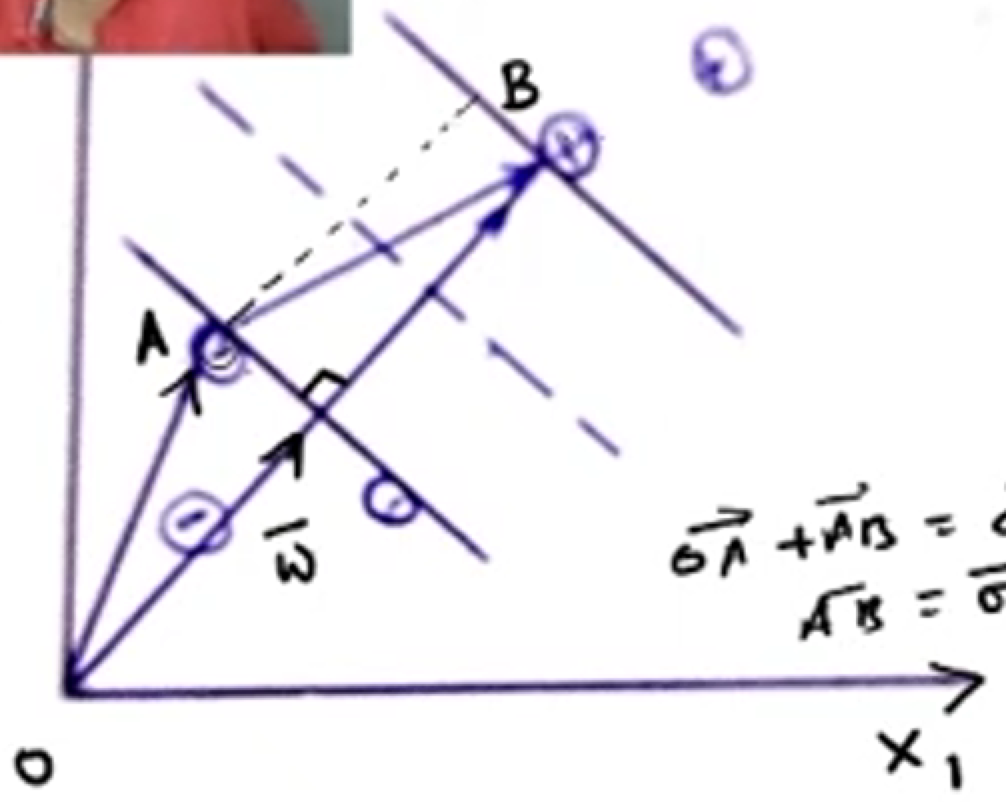
$Y^{(i)} = +1$ for +ve class

$= -1$ for -ve class

$$Y^{(i)}(\bar{w}\cdot\bar{s}^{(i)} + b) - 1 \geq 0$$

$$\boxed{Y^{(i)}(\bar{w}\cdot\bar{s}^{(i)} + b) - 1 = 0}$$ ———— (4)

(For any sample on the gutter)




$\vec{OA} + \vec{AB} = \vec{OB}$

$\vec{AB} = \vec{OB} - \vec{OA}$

3. What is the width of the gutter?

$\bar{AB} \cdot \bar{w} \rightarrow$ width of the gutter.

$$(\bar{s}_+ - \bar{s}_-) \cdot \frac{\bar{w}}{\|w\|} \quad \text{——— (5)}$$

From (4) $\Rightarrow$ For +ve Sample

$\bar{s}_+ \cdot \bar{w} + b - 1 = 0 \Rightarrow \bar{s}_+ \cdot \bar{w} = 1 - b$

For −ve Sample, $-\bar{s}_- \cdot \bar{w} - b - 1 = 0$

$\Rightarrow -\bar{s}_- \cdot \bar{w} = 1 + b$

From (5), the width of the gutter $= \frac{2}{\|w\|}$, our objective is maximize $\frac{2}{\|w\|}$

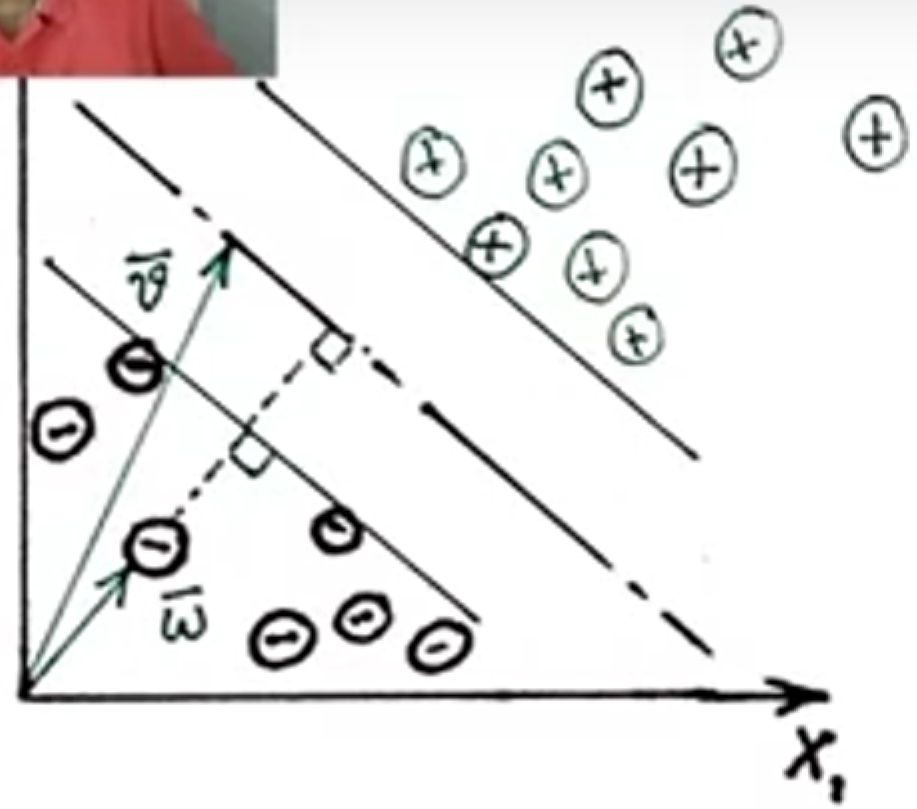
that means Minimize $\frac{1}{2}\|w\|$, that means $\underset{w}{\text{Minimize}} \ \frac{1}{2}\|w\|^2$, ——— (6)

Subject to $Y^{(i)}(\bar{w} \cdot \bar{s}^{(i)} + b) - 1 = 0$

To solve this optimization problem, we use "Lagrangian multiplier method".

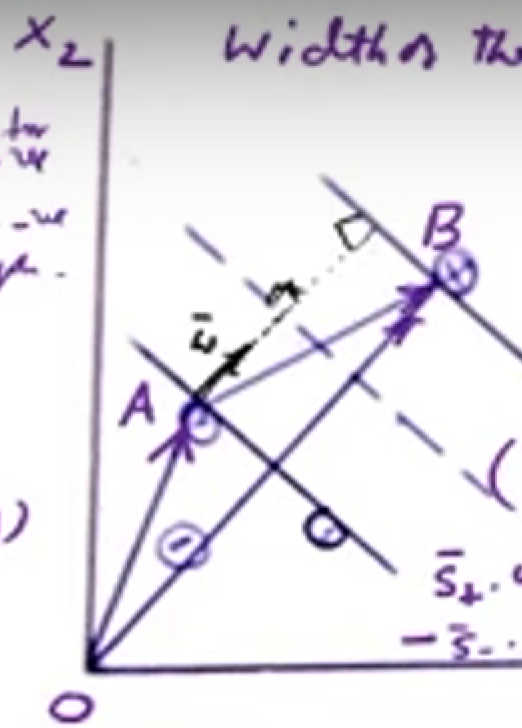
Lagrangian function $\rightarrow \quad L = \frac{1}{2}\|w\|^2 - \sum \alpha_i \left[ Y^{(i)}(\bar{w} \cdot \bar{s}^{(i)} + b) - 1 \right]$ ——— (7)

Introduction to Machine Learning, Lecture-32 ( Solving Primal Problem by Hand)




Recap:

objective function: $\min_{w} \frac{1}{2}\|w\|^2$, subject to $Y^{(i)}(\bar{w}\cdot\bar{s}^{(i)}+b)-1=0$ } $Y^{(i)}=+1$ for +ve, $=-1$ for -ve sample.




Width of the gutter:

$\bar{AB}\cdot\bar{w}=$

$\bar{OA}+\bar{AB}=\bar{OB}$

$\bar{AB}=\bar{OB}-\bar{OA}$

$(\bar{s}_+-\bar{s}_-)\cdot\frac{\bar{w}}{\|w\|}$

$\bar{s}_+\cdot w = 1-b$ } From
$-\bar{s}_-\cdot w = 1+b$ } Constraints.

We have got the expression for Lagrangian Function in the earlier lecture:

Lagrangean multiplier, $\alpha_i \geq 0$

$$L=\frac{1}{2}\|\bar{w}\|^2-\sum_{i=1}^{m}\alpha_i\left[Y^{(i)}(\bar{w}\cdot\bar{s}^{(i)}+b)-1\right] \quad \text{--- (1)}$$

$$\frac{\partial L}{\partial \bar{w}}=0 \Rightarrow \bar{w}-\sum\alpha_i Y^{(i)}\bar{s}^{(i)}=0 \quad \text{--- (2)}$$

$$\Rightarrow \boxed{\bar{w}=\sum_{i=1}^{m}\alpha_i Y^{(i)}\bar{s}^{(i)}}$$

Primal Problem Solve

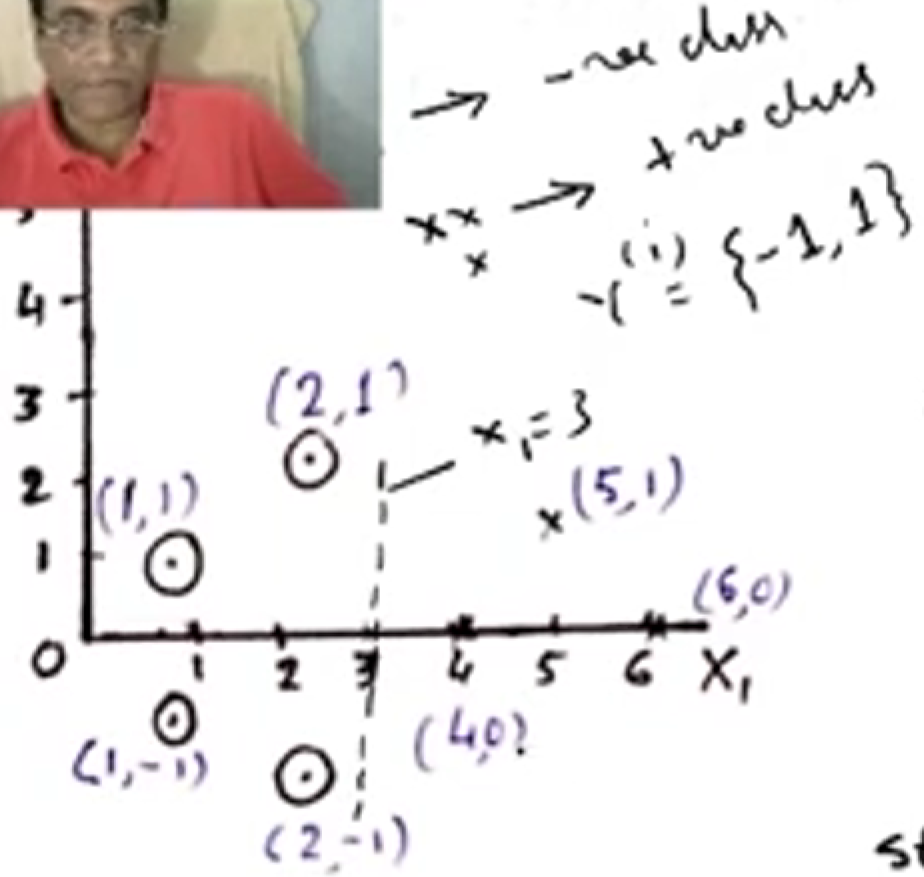
$$\frac{\partial L}{\partial b}=0 \Rightarrow -\sum\alpha_i Y^{(i)}=0 \Rightarrow \boxed{\sum\alpha_i Y^{(i)}=0} \quad \text{--- (3)}$$

putting the values of $\bar{w}$ and $b$ in (1) $\rightarrow$

$$L=\frac{1}{2}\left(\sum\alpha_i Y^{(i)}\bar{s}^{(i)}\right)^2-\sum\left(\alpha_i Y^{(i)}\cdot\bar{s}^{(i)}\right)\left(\alpha_j Y^{(j)}\cdot\bar{s}^{(j)}\right)-\underbrace{\sum\alpha_i Y^{(i)}b}_{0}+\sum\alpha_i \quad \text{--- (4)}$$

$$\Rightarrow L=\sum\alpha_i-\frac{1}{2}\sum_i\sum_j\alpha_i\alpha_j Y^{(i)}\cdot Y^{(j)}\cdot\bar{s}^{(i)}\cdot\bar{s}^{(j)}$$

5:04 / 25:18 • Software Vector Machine ›




# SVM problem solving by Hand

Let us consider two features data

Step-1, Select support vectors

$$(2,1) \;;\; (2,-1) \;;\; (4,0)$$

$$\begin{array}{ccc} \downarrow & \downarrow & \downarrow \\ \tilde{s}_1 & \tilde{s}_2 & \tilde{s}_3 \\ \alpha_1 & \alpha_2 & \alpha_3 \end{array}$$

Step-2, Augment a bias $b = 1$

$$\tilde{s}_1 = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix};\; \tilde{s}_2 = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix};\; \tilde{s}_3 = \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix}$$

Constraint Equation: $\tilde{s}^{(i)}$

$$Y^{(i)}(\bar{w}\cdot\tilde{s}^{(i)} + b) = 1$$

$$\sum_{i=1}^{3} \tilde{s}^{(i)} \sum \alpha_j Y^{(j)} \tilde{s}^{(j)} = 1 \rightarrow$$

Here, Lagrangian multipliers $\alpha_i$ $i=1,2,3$ needs to be solved using constraint equations:

$$\left.\begin{aligned} \alpha_1 \tilde{s}_1\cdot\tilde{s}_1 + \alpha_2 \tilde{s}_1\cdot\tilde{s}_2 + \alpha_3 \tilde{s}_1\cdot\tilde{s}_3 &= -1 \\ \alpha_1 \tilde{s}_2\cdot\tilde{s}_1 + \alpha_2 \tilde{s}_2\cdot\tilde{s}_2 + \alpha_3 \tilde{s}_2\cdot\tilde{s}_3 &= -1 \\ \alpha_1 \tilde{s}_3\cdot\tilde{s}_1 + \alpha_2 \tilde{s}_3\cdot\tilde{s}_2 + \alpha_3 \tilde{s}_3\cdot\tilde{s}_3 &= 1 \end{aligned}\right\}$$

$$\left.\begin{aligned} \alpha_1 \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}\begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}\begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix} &= -1 \\ \alpha_1 \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}\begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}\begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix} &= -1 \\ \alpha_1 \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix}\begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix}\cdot\begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix} &= 1 \end{aligned}\right\}$$

$$\underbrace{\begin{aligned} 6\alpha_1 + 4\alpha_2 + 9\alpha_3 &= -1 \\ 4\alpha_1 + 6\alpha_2 + 9\alpha_3 &= -1 \\ 9\alpha_1 + 9\alpha_2 + 17\alpha_3 &= 1 \end{aligned}}$$

Solve: $\alpha_1 = \alpha_2 = -3.25$

$$\alpha_3 = 3.5$$

$$\bar{w} = \sum_{i=1}^{3} \alpha_i \tilde{s}_i = \alpha_1 \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ 0 \\ -3 \end{pmatrix}$$

$$\uparrow$$

$$w^T x + b = 0$$

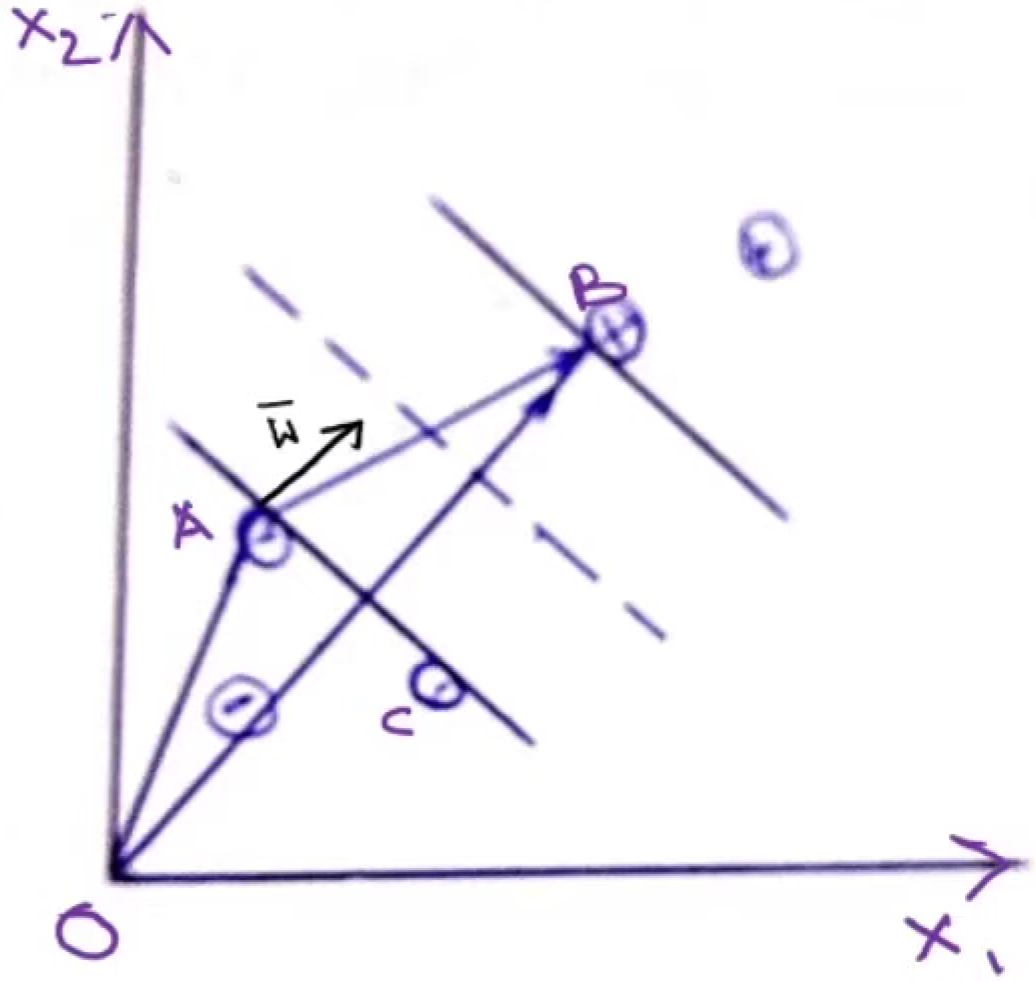
$$w_1 x_1 - 3 = 0$$

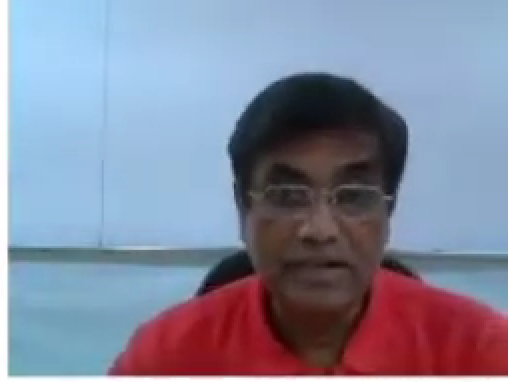
$$x_1 = 3$$

# Lagrangian Duality

- We will try to formulate the Lagrangian Dual Problem of the SVM for exploiting its capability to the fullest extent.
- It will allow us to use kernels to get optimal margin classifier even in a very higher dimensional space.
- It will allow us to derive an efficient algorithm for solving the optimization problem that will typically perform much better than the generic Q P solver.


# Developing SVM Model



Objective function $\rightarrow$ $\min\limits_{w} \frac{1}{2}\|\bar{w}\|^2$

Subject to $Y^{(i)}(\bar{w}\cdot\bar{s}^{(i)} + b) - 1 = 0$

To solve this problem let us use Lagrangian Multiplier Method.

Construct a Lagrangian function, Lagrangian multiplier.

$$L(\bar{w}, b) = \frac{1}{2}\|\bar{w}\|^2 - \sum \alpha_i \left[ Y^{(i)}(\bar{w}\cdot\bar{s}^{(i)} + b) - 1 \right] \quad \text{—— (6)}$$

How to minimize $L(w, b)$ ?

$$\frac{\partial L(w,b)}{\partial \bar{w}} = 0 \Rightarrow \bar{w} - \sum \alpha_i Y^{(i)} \cdot \bar{s}^{(i)} = 0$$

$$\Rightarrow \bar{w} = \sum \alpha_i Y^{(i)} \cdot \bar{s}^{(i)} \quad \text{—— (7)}$$

$$\frac{\partial L(\bar{w},b)}{\partial b} = 0 \Rightarrow 0 - \sum \alpha_i Y^{(i)} = 0 \Rightarrow \sum \alpha_i Y^{(i)} = 0 \quad \text{—— (8)}$$

Putting the values of $\bar{w}$ in eqn (6) $\Rightarrow$

from eq

$$L = \frac{1}{2}\left(\sum \alpha_i Y^{(i)} \bar{s}^{(i)}\right)^2 - \underbrace{\sum \left(\alpha_i Y^{(i)} \cdot \bar{s}^{(i)}\right)\left(\sum_j \alpha_j Y^{(j)} \cdot \bar{s}^{(j)}\right)}_{a} - \underbrace{\sum \alpha_i Y^{(i)} \cdots}_{=0}$$

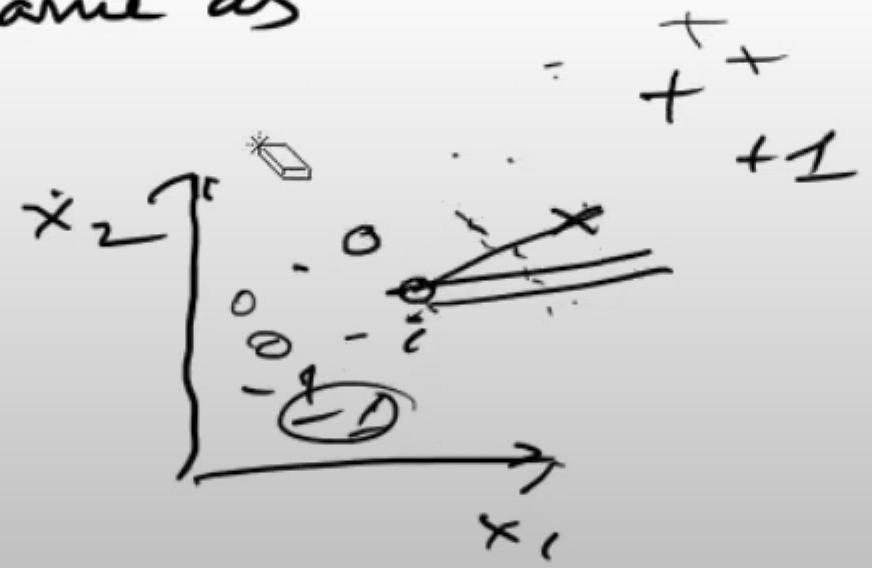
$$= \frac{1}{2}\left(\sum \alpha_i Y^{(i)} \bar{s}^{(i)}\right)\left(\sum_j \alpha_j Y^{(j)} \bar{s}^{(j)}\right) - \underbrace{\sum\left(\alpha_i Y^{(i)} \cdot \bar{s}^{(i)}\right)\left(\sum_j \alpha_j Y^{(j)} \bar{s}^{(j)}\right)}_{a} + \sum \alpha_i$$

$$L = \sum \alpha_i - \frac{1}{2}\left(\sum \alpha_i Y^{(i)} \bar{s}^{(i)}\right)\left(\sum_j \alpha_j Y^{(j)} \bar{s}^{(j)}\right) \qquad (9)$$

$$= \sum \alpha_i - \frac{1}{2}\sum_i \sum_j \alpha_i \alpha_j Y^{(i)} Y^{(j)} \underbrace{\bar{s}^{(i)} \cdot \bar{s}^{(j)}}_{\text{dot product}}$$

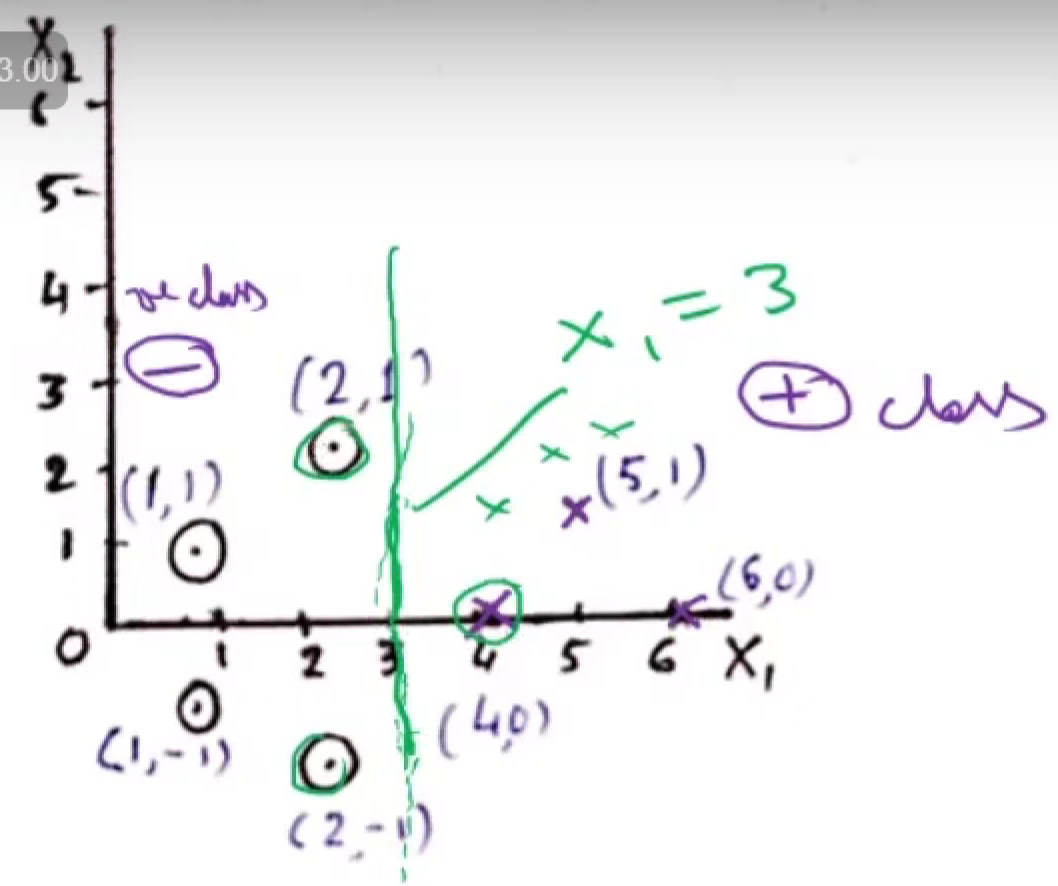
Objective is to Minimize L which same as

Maximizing the dot product



Introduction to Machine Learning, Lecture-34 ( Dual Problem for SVM)

# SVM Problem Solving by Hand




Step-1 → Select support vectors

$$\underbrace{\overset{\bar{s}_1}{(2,1)}\ ;\ \overset{\bar{s}_2}{(2,-1)}}_{-1}\ ;\ \underbrace{\overset{\bar{s}_3}{(4,0)}}_{+1} = \bar{s}_i$$

Constraint Equation $y^{(i)}(\bar{w}\,\bar{s}^{(i)} + b) = 1$ (Ref Equation No (4))

$$\tilde{s}^{(i)} = \begin{bmatrix} \bar{s}^{(i)} \\ b \end{bmatrix}$$

$$\sum_{i=1}^{3} \tilde{s}^{(i)} \sum_{j=1}^{3} \alpha_j\, y^{(j)}\, \tilde{s}^{(j)} = 1$$

$$\tilde{s}^{(1)} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix};\quad \tilde{s}^{(2)} = \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix}\quad \tilde{s}^{(3)} = \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix}$$

Step-2 → write constraint equations.

$$\tilde{s}^{(1)}(\alpha_1 \tilde{s}^{(1)} + \alpha_2 \tilde{s}^{(2)} + \alpha_3 \tilde{s}^{(3)}) = -1$$

$$\tilde{s}^{(2)}(\alpha_1 \tilde{s}^{(1)} + \alpha_2 \tilde{s}^{(2)} + \alpha_3 \tilde{s}^{(3)}) = -1$$

$$\tilde{s}^{(3)}(\alpha_1 \tilde{s}^{(1)} + \alpha_2 \tilde{s}^{(2)} + \alpha_3 \tilde{s}^{(3)}) = 1$$

Introduction to Machine Learning, Lecture-34 ( Dual Problem for SVM)

Rewriting the equations:

$$\alpha_1 \tilde{s}^{(1)} \cdot \tilde{s}^{(1)} + \alpha_2 \tilde{s}^{(1)} \cdot \tilde{s}^{(2)} + \alpha_3 \tilde{s}^{(1)} \cdot \tilde{s}^{(3)} = -1 \quad \text{—— (1)}$$

$$\alpha_1 \tilde{s}^{(2)} \cdot \tilde{s}^{(1)} + \alpha_2 \tilde{s}^{(2)} \cdot \tilde{s}^{(2)} + \alpha_3 \tilde{s}^{(2)} \cdot \tilde{s}^{(3)} = -1 \quad \text{—— (2)}$$

$$\alpha_1 \tilde{s}^{(3)} \cdot \tilde{s}^{(1)} + \alpha_2 \tilde{s}^{(3)} \cdot \tilde{s}^{(2)} + \alpha_3 \tilde{s}^{(3)} \cdot \tilde{s}^{(3)} = 1 \quad \text{—— (3)}$$

Equations (1), (2), (3) in given problem take the following form:

$$\alpha_1 \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix} = -1 \quad \text{—— (4)}$$

$$\alpha_1 \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix} = -1 \quad \text{—— (5)}$$

$$\alpha_1 \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} + \alpha_3 \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 4 \\ 0 \\ 1 \end{pmatrix} = 1 \quad \text{—— (6)}$$

$$\left.\begin{aligned} 6\alpha_1 + 4\alpha_2 + 9\alpha_3 &= -1 \quad \text{(from (4))} \\ 4\alpha_1 + 6\alpha_2 + 9\alpha_3 &= -1 \quad \text{(from (5))} \\ 9\alpha_1 + 9\alpha_2 + 17\alpha_3 &= 1 \quad \text{(from (6))} \end{aligned}\right\}$$

Solving these 3 equations we get:

$\alpha_1 = \alpha_2 = -3.25$

$\alpha_3 = 3.5$

# Formulating Dual Problem of SVM.

A software giant → wants to produce a design say, $X$.

Cost associated with that design say $f(X)$.

Primal

Minimize $f(x) = 20x_1 + 15x_2$

$x_1 \rightarrow$ office cost

$x_2 \rightarrow$ Labour

s.t.

$$\begin{cases} x_1 + x_2 \leq 60 & \text{(office cost)} \\ 2x_1 + x_2 \leq 160 & \text{(Labour cost)} \\ x_1, x_2 \geq 0 \end{cases}$$

outsource ..

↓

Dual

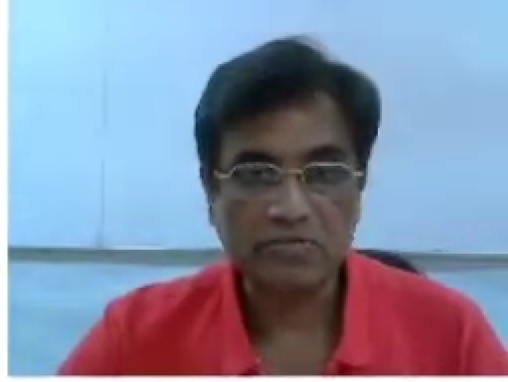
Maximize profit.

$$f(Y) = 60Y_1 + 160Y_2$$

s.t. $Y_1 + Y_2 \geq 20 \quad (x_1)$

$2Y_1 + Y_2 \geq 15 \quad (x_2)$

Max ( Min ( . . . . ) )



# Formulation of Dual Problem of SVM

- Why formulation of Dual problem is necessary?

Primal Prob : $\min_{\bar{w}} L_p(\bar{w}, \alpha, b) = \frac{1}{2}\|\bar{w}\|^2 - \sum \alpha_i \left[ Y^{(i)}(\bar{w}\bar{s}^{(i)} + b) - 1 \right]$

s.t $\quad Y^{(i)}(\bar{w}\bar{s}^{(i)} + b) - 1 \geq 0 \qquad \text{———} \quad (1)$

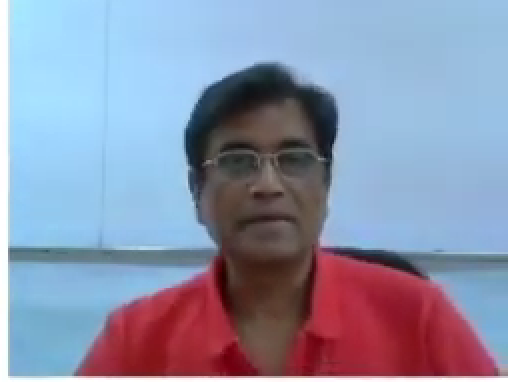
Dual Problem : $\max_{\alpha} L_D = \max_{\alpha} \left( \min_{\bar{w}} L_p(\bar{w}, b) \right)$

$$= \sum \alpha_i - \frac{1}{2} \sum_i \sum_j Y^{(i)} \cdot Y^{(j)} \cdot \alpha_i \alpha_j \langle \bar{s}^{(i)} \cdot \bar{s}^{(j)} \rangle$$

s.t $\quad \alpha_i \geq 0$

$\sum \alpha_i Y^{(i)} = 0 \qquad \text{———} \quad (2)$

$\bar{w} = \sum \alpha_i Y^{(i)} \cdot \bar{s}^{(i)}$



If eqn (2) has to be a dual of eqn (1), then there must exist

$\bar{w}^*, \alpha^*$ so that $\bar{w}^*$ is the solution to The Primal problem (Ref (1))

and $\alpha^*$ is the solution to the Dual problem.

and $p^* = D^* = L(\bar{w}^*, \alpha^*)$

But in general $D^* \leq p^*$ i.e.

$$\text{Max Min}(\text{-------}) \leq \text{Min Max}(\text{-------}) \qquad \text{———} \quad (3)$$

(3) is true for any function. Let us test it with a specific indicator function.

ex: $$\max_{y \in \{0,1\}} \underbrace{\left( \min_{x \in \{0,1\}} 1\{x = y\} \right)}_{0} \leq \min_{y \in \{0,1\}} \underbrace{\left( \max_{x \in \{0,1\}} 1\{x = y\} \right)}_{1}$$

$0 < 1$ test is a success.

Moreover, $\bar{w}^*, \alpha^*$ must satisfy K-K-T Conditions:

$\overset{1939}{\text{Karush}} - \overbrace{\text{Kuhn} - \text{Tucker}}^{1951}$

$$\frac{\partial}{\partial \bar{w}} L(\bar{w}^*, \alpha^*) = 0 \tag{4}$$

$$\alpha_i \left[ y^{(i)} \underbrace{(\bar{w} \cdot \bar{s}^{(i)} + b) - 1}_{g_i(\bar{w}^*)} \right] = 0 \tag{5}$$

$$g_i(\bar{w}^*) < 0 \tag{6}$$

$$\alpha^* \geq 0 \tag{7}$$