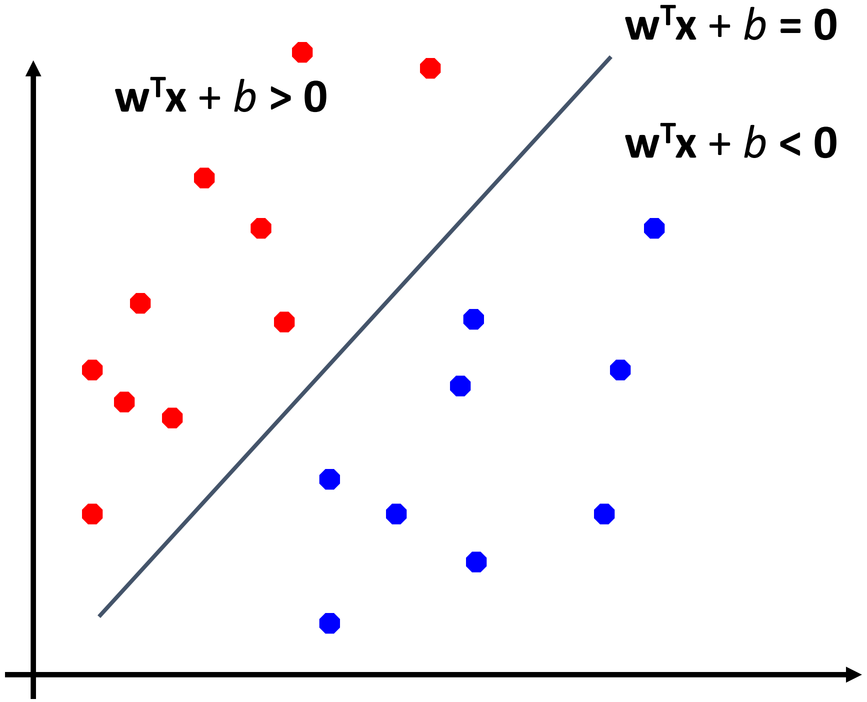
# Support Vector Machines

# Linear Separators

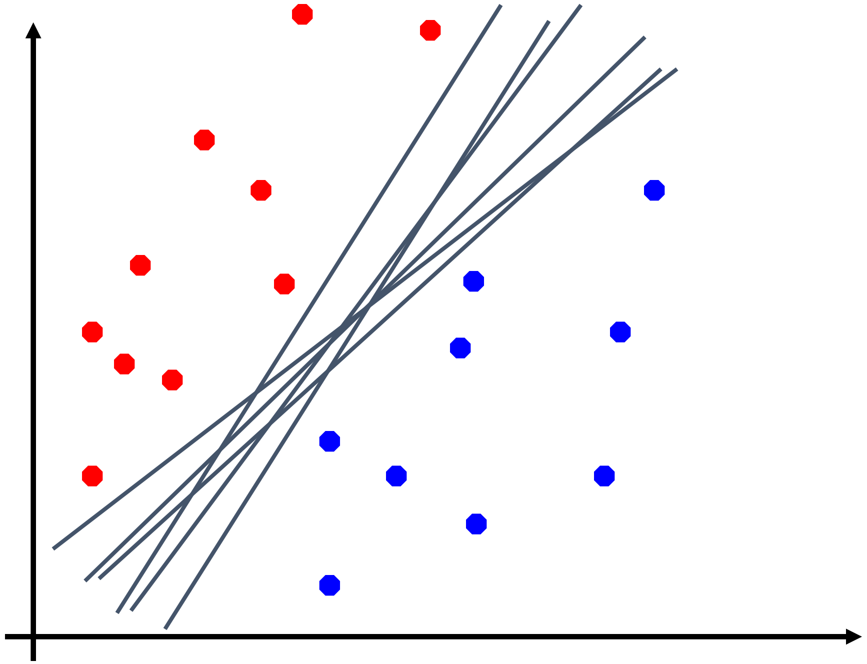
- Binary classification can be viewed as the task of separating classes in feature space:




$f(\mathbf{x}) = \text{sign}(\mathbf{w}^T\mathbf{x} + b)$

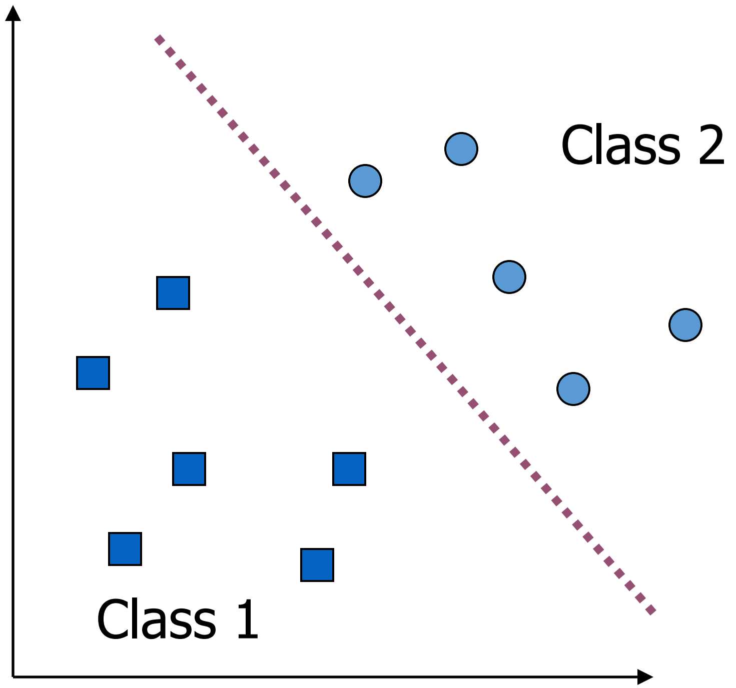
# Linear Separators

- Which of the linear separators is optimal?

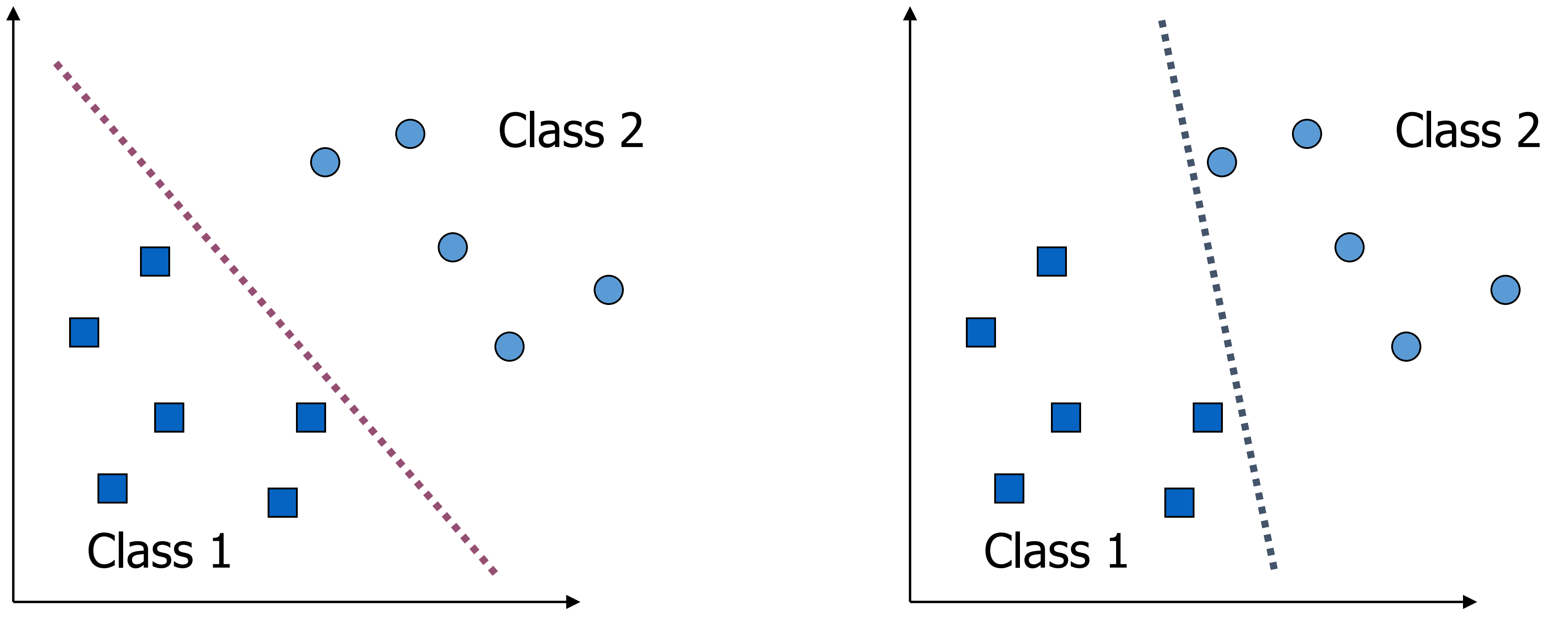


# What is a good Decision Boundary?

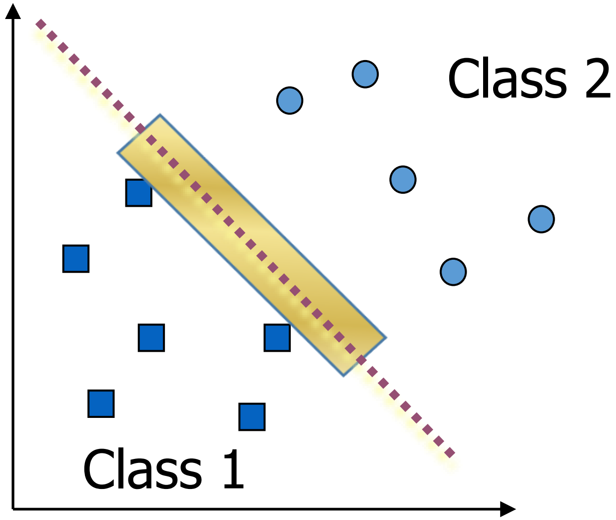
- Many decision boundaries!
  - The Perceptron algorithm can be used to find such a boundary
- Are all decision boundaries equally good?

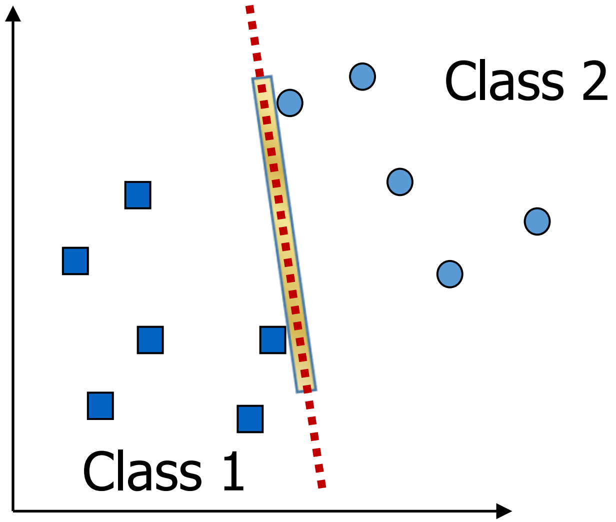


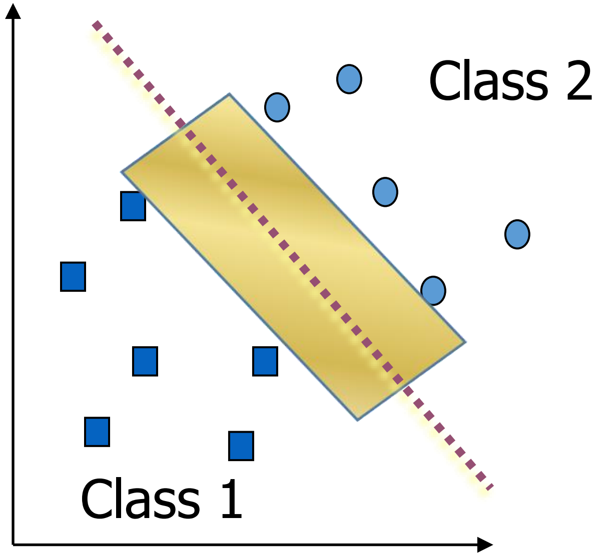
# Examples of Bad Decision Boundaries



# Better Linear Separation

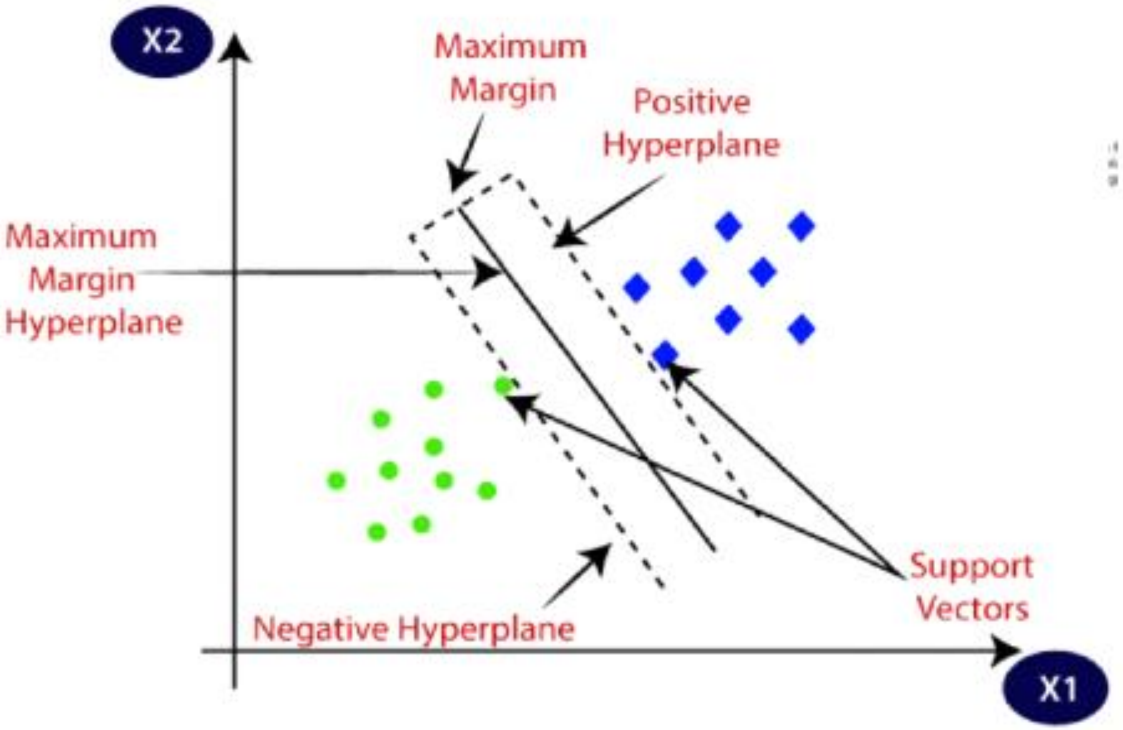


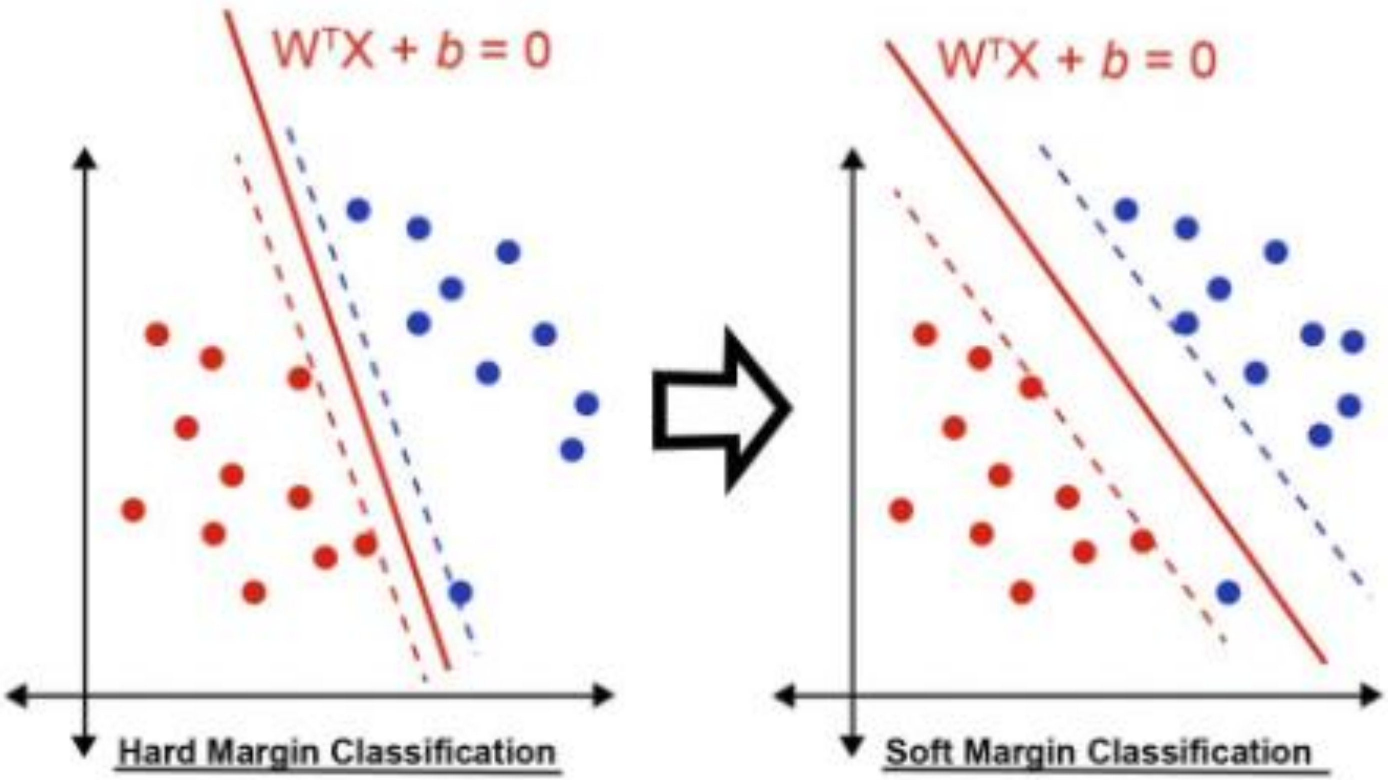




1. Why is bigger margin better?
2. Which **w** maximizes the margin?

SVM Feature




- Support Vectors
- Hyper plane
- Marginal Distance


$W^TX + b = 0$
Hard Margin Classification
$W^TX + b = 0$
Soft Margin Classification

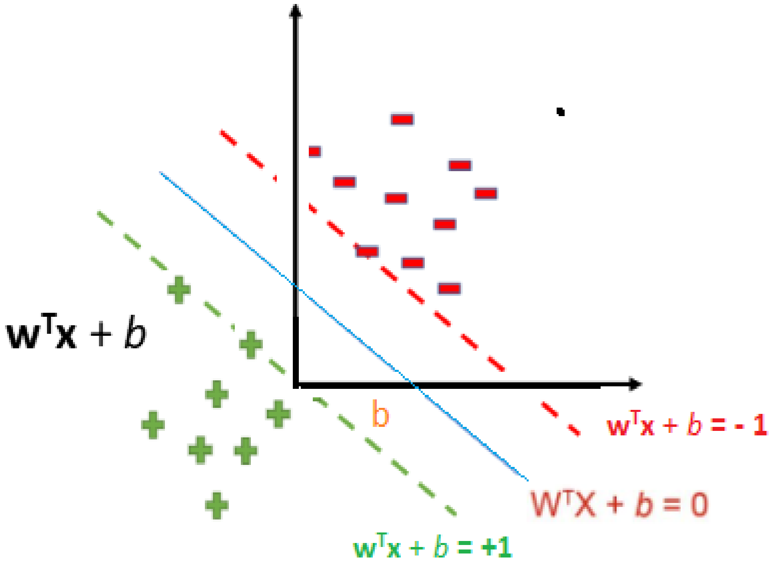
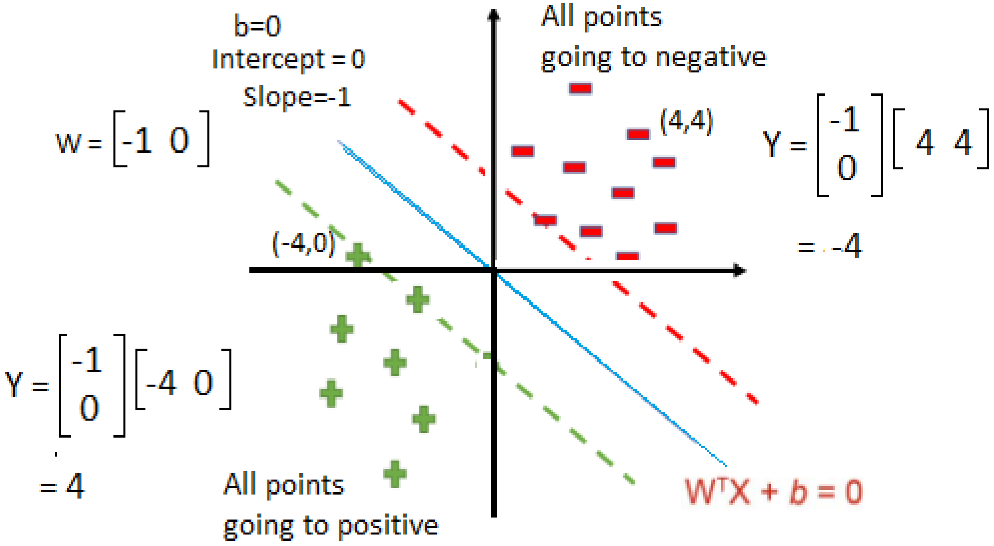



**Figure 1**




$$w^Tx_- + b = -1$$
$$w^Tx_+ + b = +1$$

--------------------

remove $W^T$ need to divided by Norm of W

$$w^T(x_+ - x_-) = 2$$

$$\frac{w^T}{||w||}(x_+ - x_-) = \frac{2}{||w||}$$

**Find out (W, b) to Max $\frac{2}{||W||}$**

$$Y_i \begin{cases} 1 & w^Tx + b >= +1 \\ -1 & w^Tx + b <= -1 \end{cases} \qquad Y_i(w^Tx + b) >= 1$$

# Finding the Decision Boundary

- The decision boundary should classify all points correctly $\Rightarrow$

$$y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1, \qquad \forall i$$

- The decision boundary can be found by solving the following constrained optimization problem

$$\text{Minimize } \frac{1}{2}||\mathbf{w}||^2$$

$$\text{subject to } y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 \qquad \forall i$$

- This is a constrained optimization problem. Solving it requires to use Lagrange multipliers

# Finding the Decision Boundary

$$\text{Minimize } \frac{1}{2}||\mathbf{w}||^2$$

$$\text{subject to } 1 - y_i(\mathbf{w}^T\mathbf{x}_i + b) \leq 0 \qquad \text{for } i = 1, \ldots, n$$

- The Lagrangian is

$$\mathcal{L} = \frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{i=1}^{n} \alpha_i \left(1 - y_i(\mathbf{w}^T\mathbf{x}_i + b)\right)$$

- $\alpha_i \geq 0$
- Note that $||\mathbf{w}||^2 = \mathbf{w}^T\mathbf{w}$

# Prerequisite

*Optimization Problems using Subject to Constraint*

$$MAX_{xy}\, Z \quad where \;\; [z = x^2 y] \qquad STC\; x^2 + \; y^2 = 1$$

*Lagrange Multiplier*

$$L(h, s, \lambda) = f(h, s) - \lambda\big(\mathrm{H(h, s)}\big)$$

*more condtions*

$$L(h, s, \lambda) = f(h, s) - \lambda 1\big(\mathrm{H1(h, s)}\big) - \lambda 2\big(\mathrm{H2(h, s)}\big)$$

***Example***

$$MAX_{hs}\, 200h^{2/3}s^{1/3}\, f(h, s)$$

$$20h + 170s = 20000\; H(h, s) \;\; [Equality\ condition]$$

$$L(h, s, \lambda) = \; 200h^{2/3}s^{1/3} - \lambda(20h + 170s - 20000)$$

$$\frac{\partial L}{\partial h} = \; 200\frac{2}{3}h^{-1/3} \; s^{1/3} - 20\lambda = 0$$

$$\frac{\partial L}{\partial s} = \; 200\frac{1}{3}h^{2/3} \; s^{-2/3} - 170\lambda = 0$$

$$\frac{\partial L}{\partial \lambda} = \; -20h - 170s + 20000 = 0$$

$$h = 666.66, s = 39.12, \lambda = 2.59$$

$$\max f(hs) = 51777$$

# Prerequisite

## Karush Kuhn Tucker

KKT Conditions

1. $Convert\ to\ Lagrange\ funtions, partially\ derive\ variables\ and\ equals\ to\ 0$
2. $\lambda_i h^i = 0$
3. $h^i \geq 0$
4. $\lambda_i \geq 0$

$$Max\ -x_1{}^2 - x_2{}^2 - x_3{}^2 + 4x_1 + 6x_2$$

$$STC\ \ x_1 + x_2 \leq 2$$

$$2x_1 + 3x_2 \leq 12$$

$$x_1, x_2 \geq 0$$

Conditions 1:

$$\mathrm{L}(x_1, x_2, x_3, \lambda_1, \lambda_2) = -x_1{}^2 - x_2{}^2 - x_3{}^2 + 4x_1 + 6x_2 - \lambda_1(x_1 + x_2 - 2) - \lambda_2(2x_1 + 3x_2 - 12)$$

$$\frac{\partial L}{\partial x_1} = 2x_1 + 4 - \lambda_1 - 2\lambda_2 = 0 \ \ \ldots..(1a)$$

$$\frac{\partial L}{\partial x_2} = 2x_2 + 6 - \lambda_1 - 3\lambda_2 = 0 \ \ \ldots\ldots(1b)$$

$$\frac{\partial L}{\partial x_3} = 2x_3 = 0 \quad i.e:\ x_3 = 0$$

Conditions 2:

$$\lambda_1(x_1 + x_2 - 2) = 0 \ \ \ldots\ldots(2a)$$

$$\lambda_2(2x_1 + 3x_2 - 12) = 0 \ \ \ldots\ldots(2b)$$

Conditions 3:

$$x_1 + x_2 - 2 \ \leq 0 \ \ \ldots\ldots(3a)$$

$$2x_1 + 3x_2 - 12 \ \leq 0 \ \ \ldots..(3b)$$

Conditions 4:

$$\lambda_1 \geq 0, \lambda_2 \geq 0$$

# Prerequisite

## Karush Kuhn Tucker

Case 1: $\lambda_1 = 0\,,\ \lambda_2 = 0$

Substitute 1a, 1b → $x_1 = 2\,,\ x_2 = 3$

Substitute $x_1\ x_2\ in\ 3a, 3b$ $\quad x_1 + x_2 - 2 \le 0$

$5 - 2 \le 0$

$3 \le 0$ X

$2x_1 + 3x_2 - 12 \le 0$

$1 \le 0$ X

Case 2: $\lambda_1 \ne 0\,,\ \lambda_2 \ne 0$

Means from condition 2

$x_1 + x_2 - 2 = 0, 2x_1 + 3x_2 - 12 = 0\ \ by\ solving\ \ x_2 = 8\,, x_1 = -6$

Substitute in 1a, 1b → Solve $\lambda_1\,,\ \lambda_2$

$\lambda_2 = -26$ X

Case 3: $\lambda_1 = 0\,,\ \lambda_2 \ne 0$

Substitute in 1a, 1b

$-2x_1 + 4 - 2\lambda_2 = 0$

$-2x_1 + 6 - 3\lambda_2 = 0$ , solving $x_1 = \frac{2}{3}x_2$

$\lambda_2 \ne 0\ , so$

$2x_1 + 3x_2 - 12 = 0$

$\frac{4}{3}x_1 + 3x_2 - 12 = 0$

$x_1 = 2\,,\ x_2 = 3$

$x_1 + x_2 - 2 \le 0$

$5 - 2 \le 0$ X

$2x_1 + 3x_2 - 12 \le 0$

$4 + 9 - 12 \le 0$ X

Case 4: $\lambda_1 \ne 0\,,\ \lambda_2 = 0$

$\lambda_1 = 3\,,\ \lambda_2 = 0\ \ ,\ x_1 = \frac{1}{2}\ , x_2 = {}^{3}/_{2}$ ✓

$x_1 + x_2 - 2 \le 0\ \ (0 \le 0)\ \ \ 2x_1 + 3x_2 - 12\ (-13 \le 0)$

# Prerequisite

**Primal and dual problem for understanding support vector machine:**

$M_{in} f(w)$

STC $g_i(w) \leq 0 \quad i = 1 \ldots \ldots k$

$h_i(w) = 0 \quad i = 1 \ldots \ldots l$

Generalized Lagrange function:

L $(w, \alpha, \beta) \; = f(w) + \sum_{i=1}^{k} \alpha_i g_i(w) + \sum_{i=1}^{l} \beta_i h_i(w)$

Define: $\theta_p(w) = \; Max_{\alpha\, \beta, \alpha \leq 0}\, L(w, \alpha, \beta)$

$\theta_p(w) = Max_{\alpha\, \beta, \alpha \leq 0}\, f(w) + \sum_{i=1}^{k} \alpha_i g_i(w) + \sum_{i=1}^{l} \beta_i h_i(w)$

If $g_i(w) > 0$ $[violates\ condition]$ $\qquad \theta_p(w) = \infty$

If $h_i(w) \neq 0$ $[violates\ condition]$ $\qquad \theta_p(w) = \infty$

If $g_i(w) \,.\, h_i(w)$ $[satisfies\ condition]$ $\theta_p(w) = f(w)$

So, $\quad \theta_p(w) = \begin{cases} f(w) \rightarrow satisfies \\ \infty \rightarrow violates \end{cases}$

Primal problem:

$p^* = min_w \; \theta_p(w)$

$p^* = \; min_w Max_{\alpha\, \beta, \alpha \leq 0}$ L$(w, \alpha, \beta)$

Dual problem:

$d^* = Max_{\alpha\, \beta, \alpha \leq 0}\, min_w$ L $(w, \alpha, \beta)$

$= Max_{\alpha\, \beta, \alpha \leq 0} \;\; \theta_d(\alpha, \beta)$

$d^* \;\leq\; p^*$ But under some conditions $d^* = p^*$

$\exists \; w^* \alpha^* \beta^*$

Where $w^*$ solution to Primal,

$\alpha^* \beta^*$ Solution to Dual,

$d^* = p^*$ ,

$w^* \alpha^* \beta^*$ Satisfy KKT conditions,

1) Derivative w.r.t variable =0
2) $\alpha_i g_i(w) = 0$
3) $g_i(w) \leq 0$
4) $\alpha_i \geq 0$

Fact: $MaxMinf$(x) $\leq MinMaxf(x) \quad Example:\; MaxMin\ Sin(x+y) \leq MinMaxSin(x+y)$

# Gradient with respect to *w* and *b*

- Setting the gradient of $w_{\mathcal{L}}$ :. **w** and b to zero, we have

$$L = \frac{1}{2} w^T w + \sum_{i=1}^{n} \alpha_i \left(1 - y_i \left(w^T x_i + b\right)\right) =$$

$$= \frac{1}{2} \sum_{k=1}^{m} w^k w^k + \sum_{i=1}^{n} \alpha_i \left( 1 - y_i \left( \sum_{k=1}^{m} w^k x_i^{\,k} + b \right) \right)$$

n: no of examples, m: dimension of the space

$$\begin{cases} \dfrac{\partial L}{\partial w^k} = 0, \forall k \\ \dfrac{\partial L}{\partial b} = 0 \end{cases}$$

$$\mathbf{w} + \sum_{i=1}^{n} \alpha_i (-y_i) \mathbf{x}_i = \mathbf{0} \quad \Rightarrow \quad \mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i$$

$$\sum_{i=1}^{n} \alpha_i y_i = 0$$

# The Dual Problem

- If we substitute $\mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i$, we have $\mathcal{L}$

$$
\begin{aligned}
\mathcal{L} &= \frac{1}{2} \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i^T \sum_{j=1}^{n} \alpha_j y_j \mathbf{x}_j + \sum_{i=1}^{n} \alpha_i \left( 1 - y_i \Big( \sum_{j=1}^{n} \alpha_j y_j \mathbf{x}_j^T \mathbf{x}_i + b \Big) \right) \\
&= \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j + \sum_{i=1}^{n} \alpha_i - \sum_{i=1}^{n} \alpha_i y_i \sum_{j=1}^{n} \alpha_j y_j \mathbf{x}_j^T \mathbf{x}_i - b \sum_{i=1}^{n} \alpha_i y_i \\
&= -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j + \sum_{i=1}^{n} \alpha_i
\end{aligned}
$$

Since

$$\sum_{i=1}^{n} \alpha_i y_i = 0$$

- This is a function of $\alpha_i$ only

# The Dual Problem

- The new objective function is in terms of $\alpha_i$ only
- It is known as the dual problem: if we know **w**, we know all $\alpha_i$; if we know all $\alpha_i$, we know **w**
- The original problem is known as the primal problem
- The objective function of the dual problem needs to be maximized (comes out from the KKT theory)
- The dual problem is therefore:

$$\text{max. } W(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1,j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

$$\text{subject to } \alpha_i \geq 0, \qquad \sum_{i=1}^{n} \alpha_i y_i = 0$$

Properties of $\alpha_i$ when we introduce the Lagrange multipliers

The result when we differentiate the original Lagrangian w.r.t. b

# The Dual Problem

$$\text{max. } W(\boldsymbol{\alpha}) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1,j=1}^{n} \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

$$\text{subject to } \alpha_i \geq 0, \sum_{i=1}^{n} \alpha_i y_i = 0$$

- This is a quadratic programming (QP) problem
  - A global maximum of $\alpha_i$ can always be found
- **w** can be recovered by

$$\mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i$$

# The Solution - QP

$$\underset{\alpha}{\text{Max}} \quad \underbrace{(-1)^T}_{\text{linear}} \alpha \; - \tfrac{1}{2}\, \alpha^T \underbrace{\begin{pmatrix} y_1 y_1 \mathbf{x_1 x_1} & y_1 y_2 \mathbf{x_1 x_2} & \dots & y_1 y_N \mathbf{x_1 x_N} \\ y_2 y_1 \mathbf{x_2 x_1} & y_2 y_2 \mathbf{x_2 x_2} & \dots & y_2 y_N \mathbf{x_2 x_N} \\ \dots & \dots & \dots & \dots \\ y_N y_1 \mathbf{x_N x_1} & y_N y_2 \mathbf{x_N x_2} & \dots & y_N y_N \mathbf{x_N x_N} \end{pmatrix}}_{\text{quadratic coefficients}} \alpha$$

Subject to $\quad y^T \alpha = 0$

$$0 \le \alpha \le \infty$$

# QP Solver provides us α

Solution: $\boldsymbol{\alpha} = \alpha_1, \alpha_2, \ldots, \alpha_N$

- Note: **w** need not be formed explicitly

$$\mathbf{w} = \sum_{n=1}^{N} \alpha_n y_n \mathbf{x}_n$$

**KKT** Condition: For n = 1, 2, …, N

$$\alpha_i\left(1 - y_i\left(w^T x_i + b\right)\right) = 0$$

$\alpha_n > 0 \Rightarrow \mathbf{x_n}$ is **support vector**

# A Geometrical Interpretation