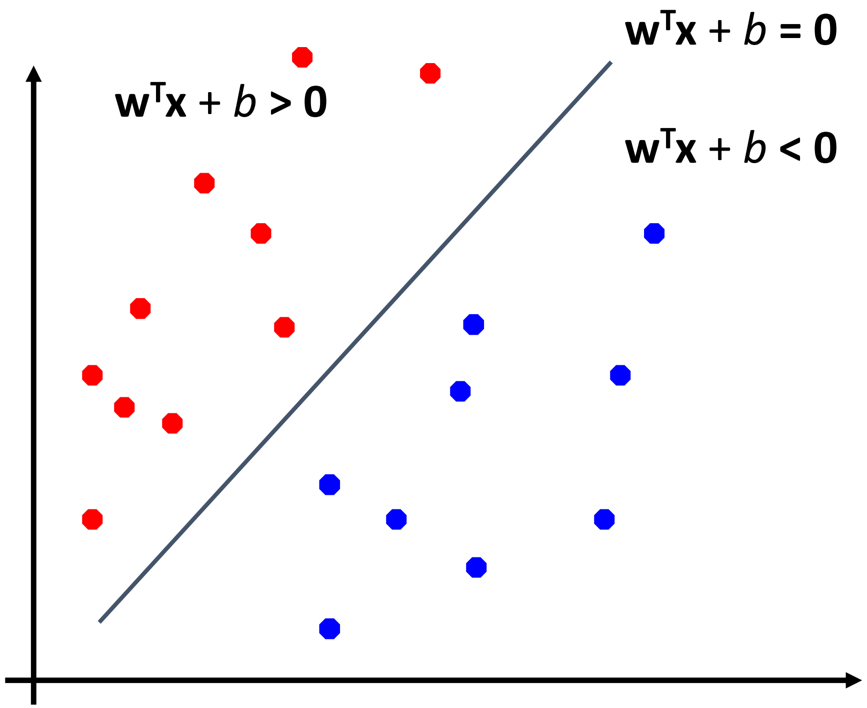
SVM

Support Vector Machines

Linear Separators

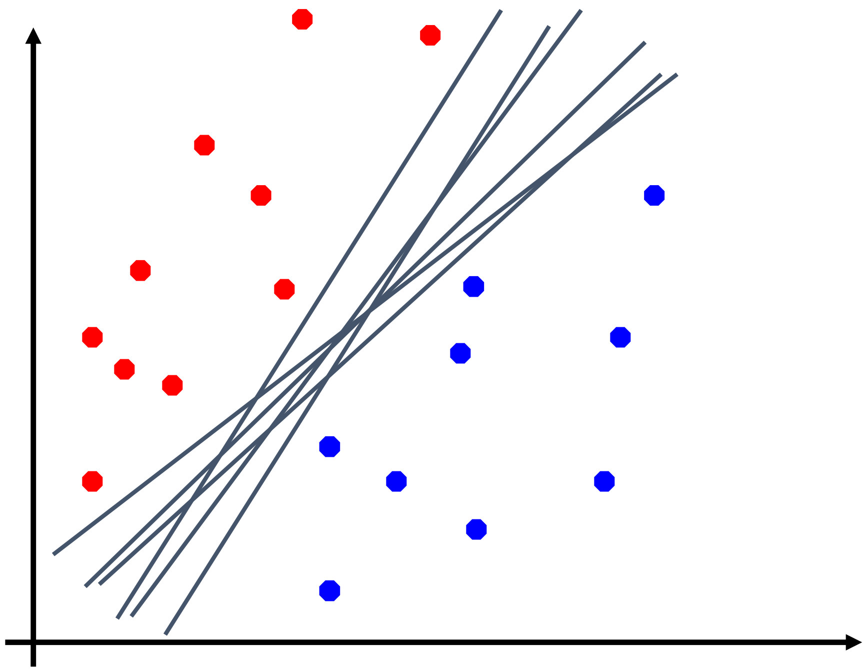
- Binary classification can be viewed as the task of separating classes in feature space:



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

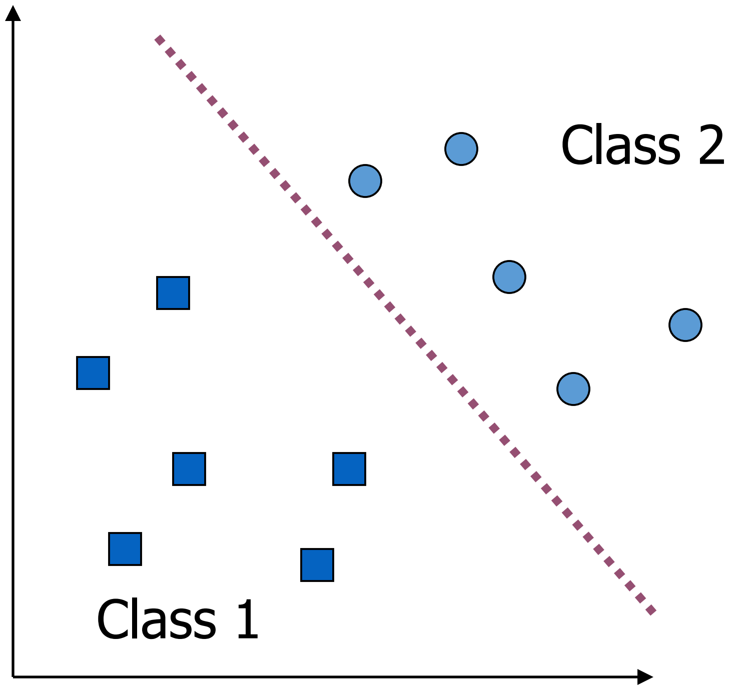
Linear Separators

- Which of the linear separators is optimal?

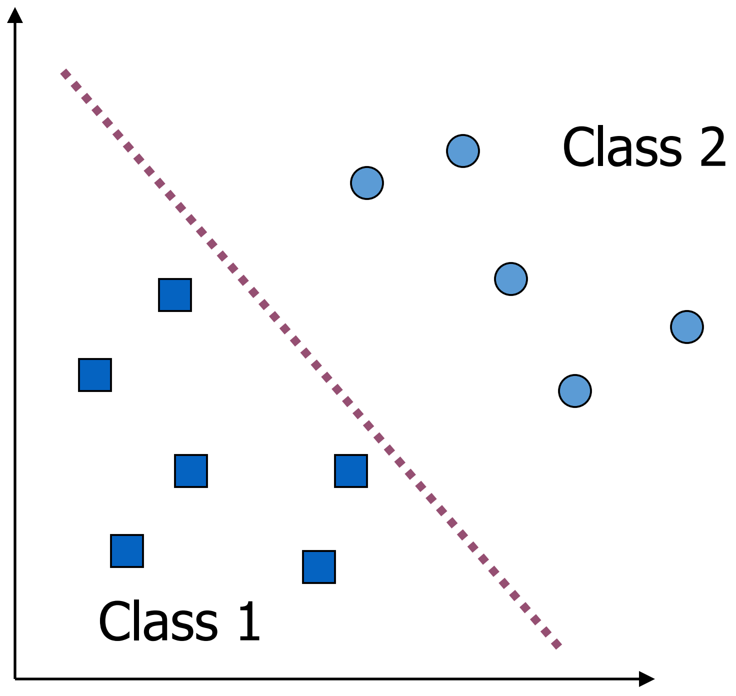


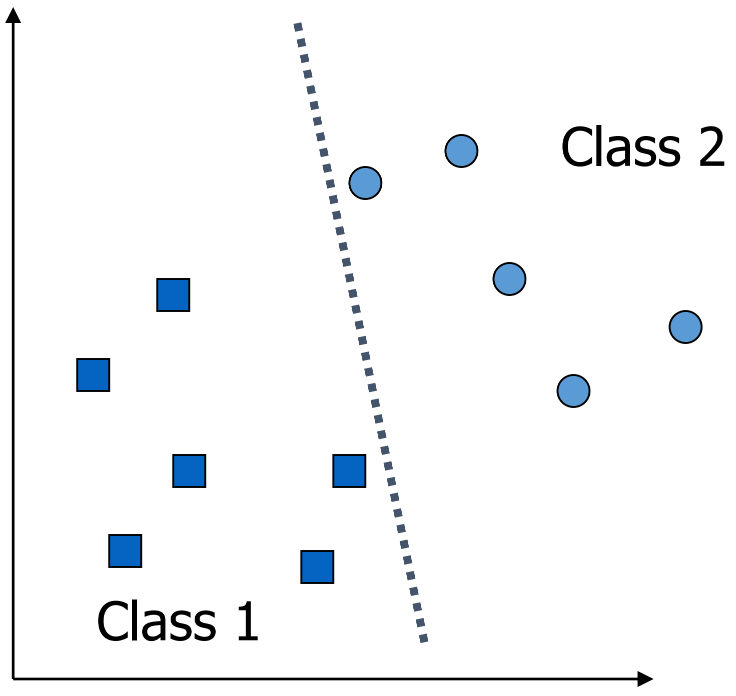
What is a good Decision Boundary?

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

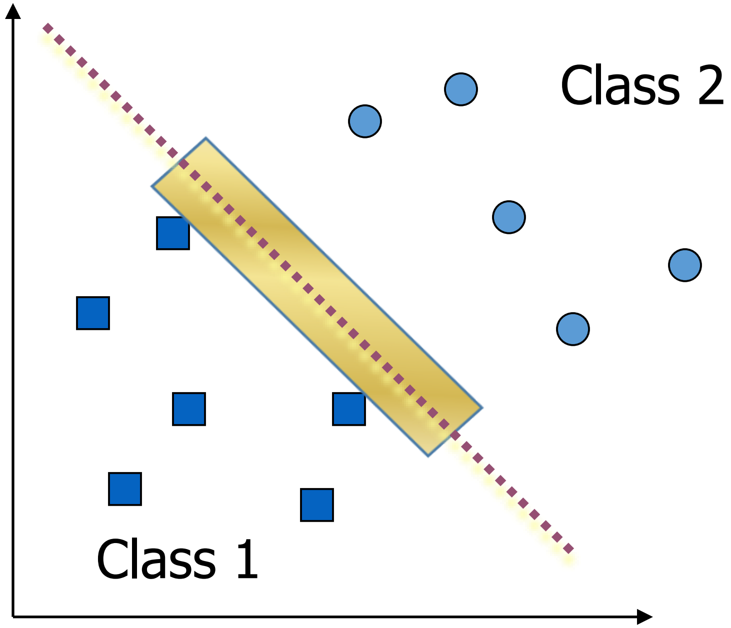


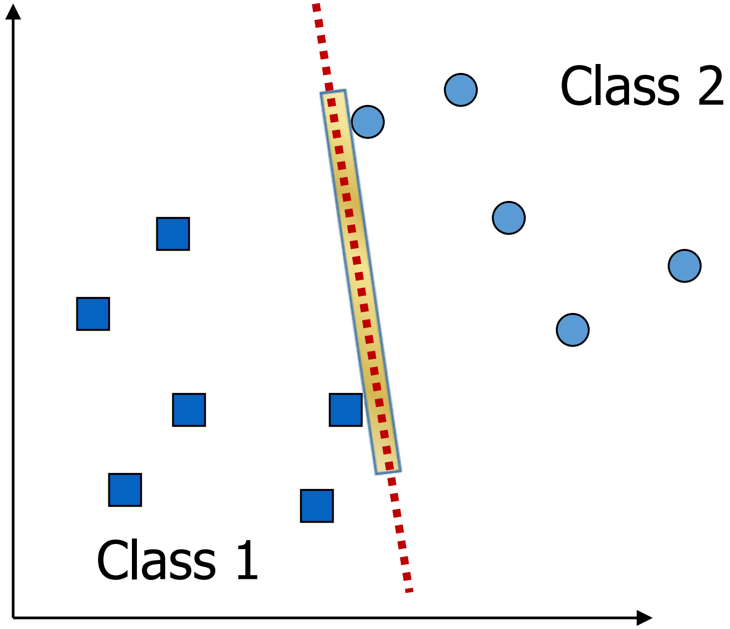
Examples of Bad Decision Boundaries

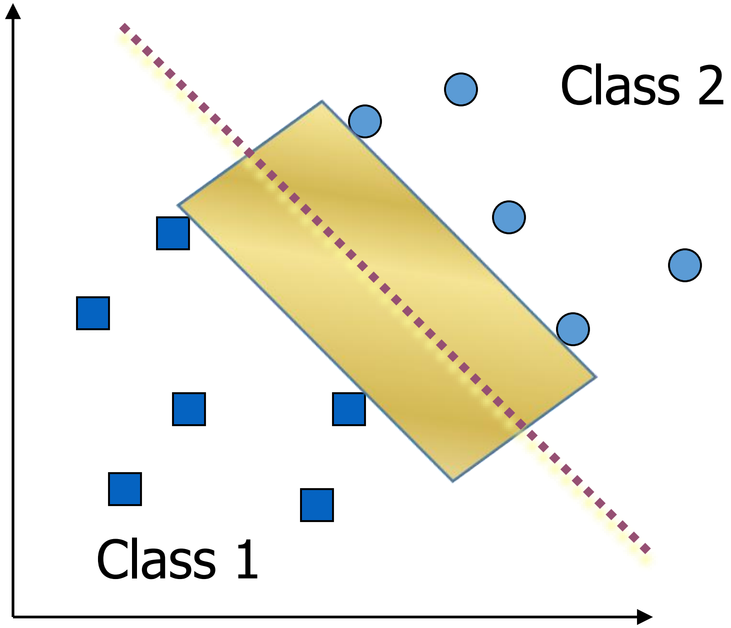




Better Linear Separation

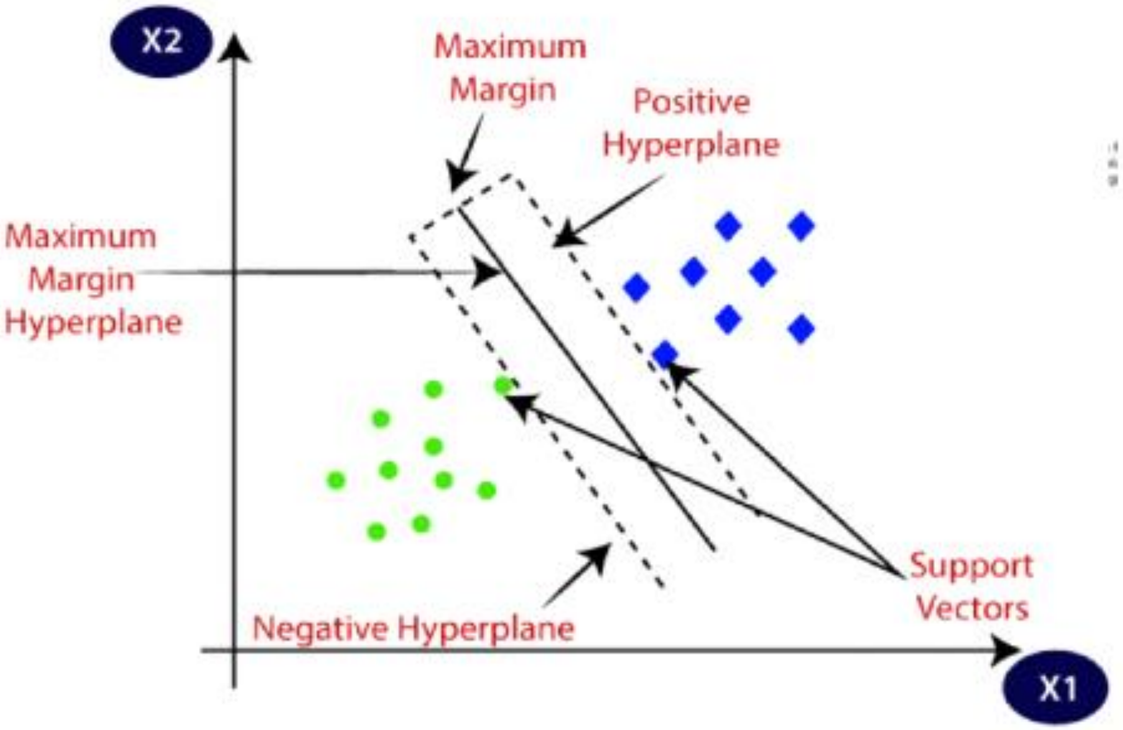




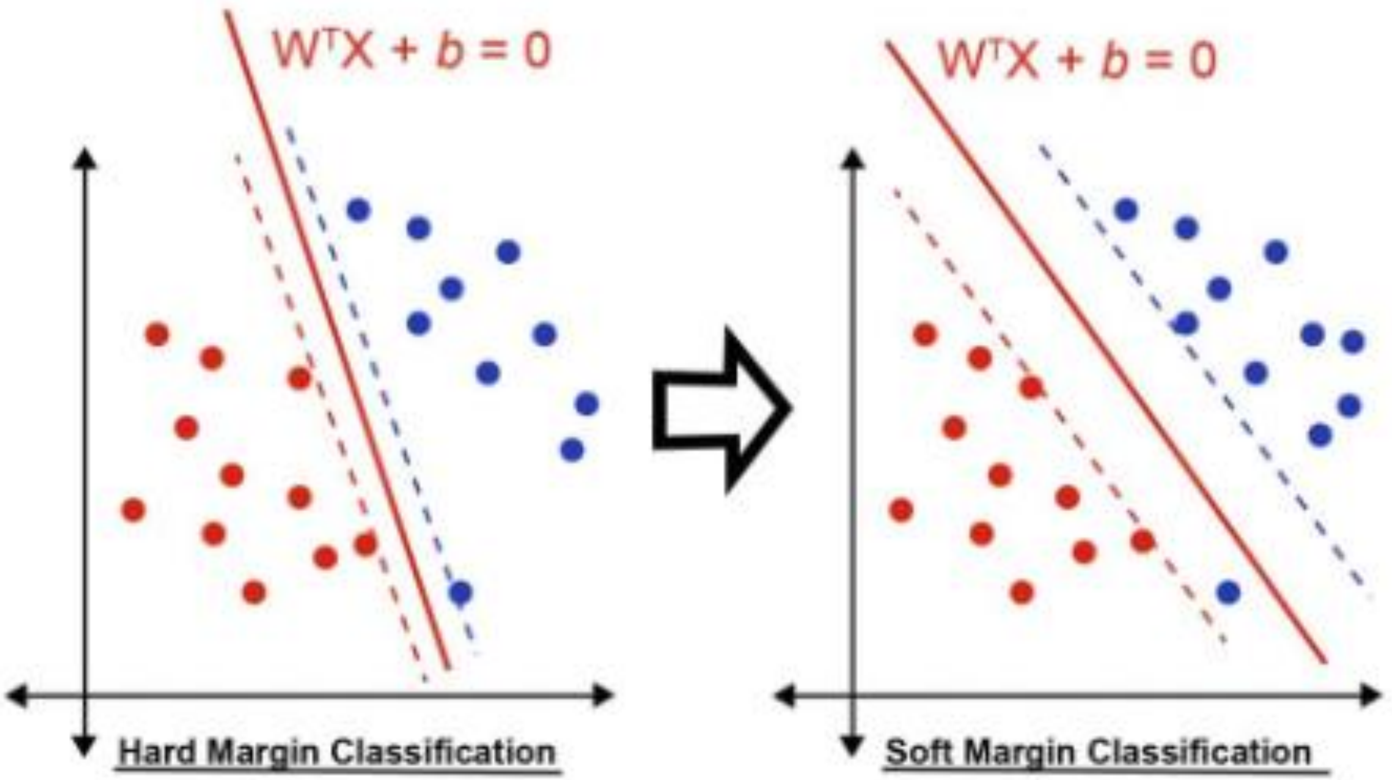


1. Why is bigger margin better?
2. Which **w** maximizes the margin?

SVM Feature



- Support Vectors
- Hyper plane
- Marginal Distance


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

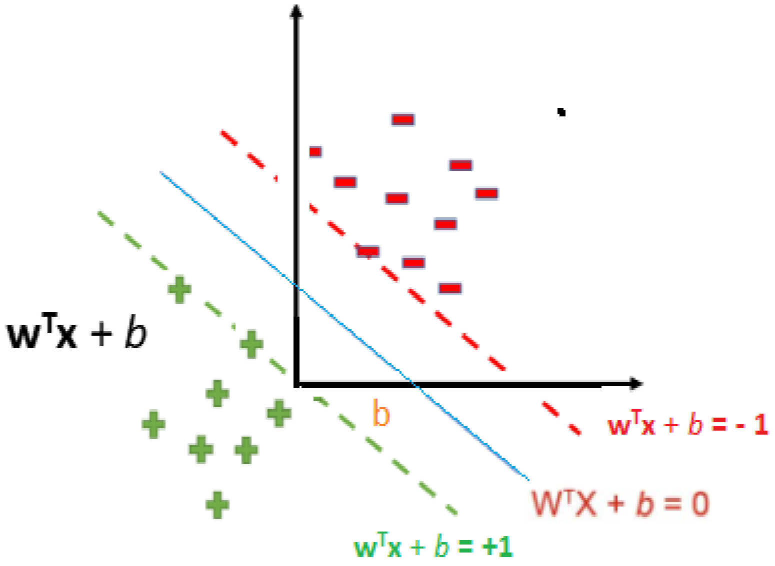
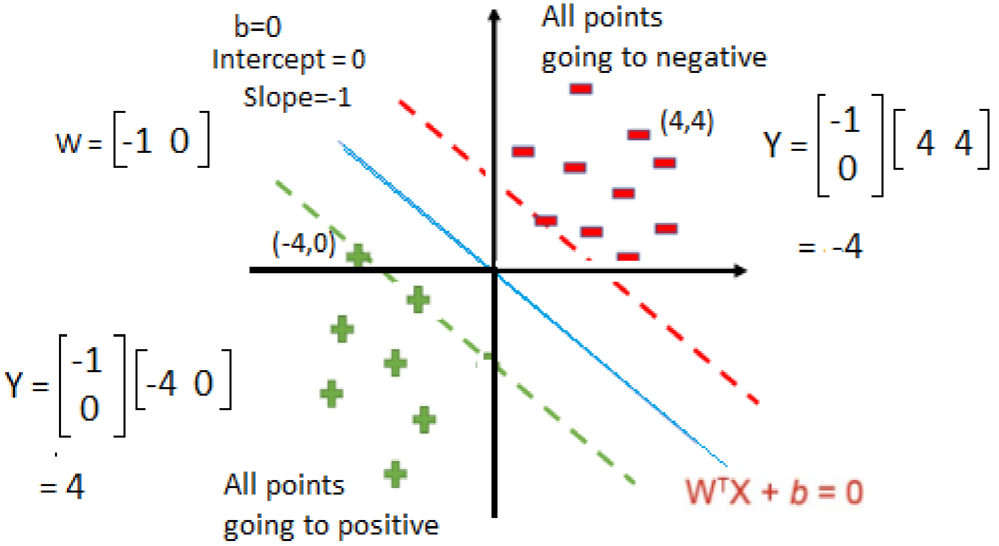


Figure 1



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

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

$Optimization\ Problems\ using\ Subject\ to\ Constraint$

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

$Lagrange\ Multiplier$

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

$more\ condtions$

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

$\boldsymbol{Example}$

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

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

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

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

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

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

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

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

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:

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

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

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

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

Conditions 2:

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

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

Conditions 3:

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

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

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: $\lambda_1 \neq 0$, $\lambda_2 \neq 0$

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 λ_1 , λ_2

$\lambda_2 = -26$ X

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

Substitute in 1a, 1b

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

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

$\lambda_2 \neq 0\ , so$

$2x_1 + 3x_2 - 12 = 0$

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

$x_1 = 2$, $x_2 = 3$

$x_1 + x_2 - 2 \le 0$

$5 - 2 \le 0$ X

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

$4 + 9 - 12 \le 0$ X

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

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

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

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

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

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

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

Primal problem:

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

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

Dual problem:

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

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

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

$d^* \leq p^*$ 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) $\alpha_i \geq 0$

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

$$\begin{cases} \dfrac{\partial L}{\partial w^k} = 0, \forall k \\ \dfrac{\partial L}{\partial b} = 0 \end{cases} \qquad \begin{array}{c} \mathbf{w} + \sum_{i=1}^{n} \alpha_i (-y_i) \mathbf{x}_i = \mathbf{0} \quad \Rightarrow \quad \mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i \\ \sum_{i=1}^{n} \alpha_i y_i = 0 \end{array}$$

The Dual Problem

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

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

Since

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

- This is a function of α_i only

The Dual Problem

- The new objective function is in terms of α_i only
- It is known as the dual problem: if we know **w**, we know all α_i; if we know all α_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 α_i when we introduce the Lagrange multipliers

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

The Dual Problem

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

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

- This is a quadratic programming (QP) problem
  - A global maximum of α_i can always be found

- **w** can be recovered by

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

SVM Classifier to Find Hyperplane – Solved Example

X1	X2	Class
2	2	-1 √
4	5	+1
7	4	+1

- $N = 3$
- $\vec{x}_1 = (2, 2)$
- $\vec{x}_2 = (4, 5)$
- $\vec{x}_3 = (7, 4)$
- $y_1 = -1$
- $y_2 = +1$
- $y_3 = +1$

$$f(\vec{x}) = \vec{w} \cdot \vec{x} - b$$

- $\vec{\alpha} = (\alpha_1, \alpha_2, \alpha_3)$
- subject to the conditions
- $\sum_{i=1}^{N} \alpha_i y_i = -\alpha_1 + \alpha_2 + \alpha_3 = 0$
- $\alpha_1 > 0, \alpha_2 > 0, \alpha_3 > 0$

SVM Classifier to Find Hyperplane – Solved Example – Step 1

$$\phi(\vec{\alpha}) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1,j=1}^{N} \alpha_i \alpha_j y_i y_j (\vec{x}_i \cdot \vec{x}_j)$$

$$= \sum_{i=1}^{3} \alpha_i - \frac{1}{2} \sum_{i=1,j=1}^{3} \alpha_i \alpha_j y_i y_j (\vec{x}_i \cdot \vec{x}_j)$$

$$(\vec{x}_1 \cdot \vec{x}_1) = 08, \quad (\vec{x}_1 \cdot \vec{x}_2) = 18, \quad (\vec{x}_1 \cdot \vec{x}_3) = 22$$

$$(\vec{x}_2 \cdot \vec{x}_1) = 18, \quad (\vec{x}_2 \cdot \vec{x}_2) = 41, \quad (\vec{x}_2 \cdot \vec{x}_3) = 48,$$

$$(\vec{x}_3 \cdot \vec{x}_1) = 22, \quad (\vec{x}_3 \cdot \vec{x}_2) = 48, \quad (\vec{x}_3 \cdot \vec{x}_3) = 65$$

$$\phi(\vec{\alpha}) = (\alpha_1 + \alpha_2 + \alpha_3) - \frac{1}{2}[8\alpha_1^2 + 41\alpha_2^2 + 65\alpha_3^2 - 36\alpha_1\alpha_2 - 44\alpha_1\alpha_3 + 96\alpha_2\alpha_3]$$

$$\phi(\vec{\alpha}) = 2(\alpha_2 + \alpha_3) - \frac{1}{2}(13\alpha_2^2 + 32\alpha_2\alpha_3 + 29\alpha_3^2)$$

$$N = 3$$

$$\vec{x}_1 = (2, 2)$$

$$\vec{x}_2 = (4, 5)$$

$$\vec{x}_3 = (7, 4)$$

$$y_1 = -1$$

$$y_2 = +1$$

$$y_3 = +1$$

$$-\alpha_1 + \alpha_2 + \alpha_3 = 0$$

SVM Classifier to Find Hyperplane – Solved Example – Step 1

- Find values of α_1, α_2 and α_3 which maximizes

$$\phi(\vec{\alpha}) = 2(\alpha_2 + \alpha_3) - \frac{1}{2}(13\alpha_2^2 + 32\alpha_2\alpha_3 + 29\alpha_3^2)$$

- For $\phi(\vec{\alpha})$ to be maximum we must have

$$\frac{\partial \phi}{\partial \alpha_2} = 0, \quad \frac{\partial \phi}{\partial \alpha_3} = 0$$

- That is,

$$2 - 13\alpha_2 - 16\alpha_3 = 0, \quad 2 - 16\alpha_2 - 29\alpha_3 = 0$$

- Solving these, we get

$$\alpha_2 = \frac{26}{121}, \quad \alpha_3 = -\frac{6}{121} \quad \alpha_1 = \frac{20}{121}$$

$$N = 3$$

$$\vec{x}_1 = (2,2)$$

$$\vec{x}_2 = (4,5)$$

$$\vec{x}_3 = (7,4)$$

$$y_1 = -1$$

$$y_2 = +1$$

$$y_3 = +1$$

$$-\alpha_1 + \alpha_2 + \alpha_3 = 0$$

SVM Classifier to Find Hyperplane – Solved Example – Step 2

$$\vec{w} = \sum_{i=1}^{N} \alpha_i y_i \vec{x}_i$$

$$= \frac{20}{121}(-1)(2,2) + \frac{26}{121}(+1)(4,5) - \frac{6}{121}(+1)(7,4)$$

$$= \left(\frac{2}{11}, \frac{6}{11}\right)$$

$$\alpha_1 = \frac{20}{121}$$

$$\alpha_2 = \frac{26}{121}$$

$$\alpha_3 = -\frac{6}{121}$$

$$N = 3$$

$$\vec{x}_1 = (2,2)$$

$$\vec{x}_2 = (4,5)$$

$$\vec{x}_3 = (7,4)$$

$$y_1 = -1$$

$$y_2 = +1$$

$$y_3 = +1$$

$$-\alpha_1 + \alpha_2 + \alpha_3 = 0$$

SVM Classifier to Find Hyperplane – Solved Example – Step 3

$$b = \frac{1}{2}\left(\min_{i:y_i=+1}(\vec{w}\cdot\vec{x}_i) + \max_{i:y_i=-1}(\vec{w}\cdot\vec{x}_i)\right)$$

$$= \frac{1}{2}\left(\min\{(\vec{w}\cdot\vec{x}_2),(\vec{w}\cdot\vec{x}_3)\} + \max\{(\vec{w}\cdot\vec{x}_1)\}\right)$$

$$= \frac{1}{2}\left(\min\{\tfrac{38}{11},\tfrac{38}{11}\} + \max\{\tfrac{16}{11}\}\right)$$

$$= \frac{1}{2}\left(\frac{38}{11} + \frac{16}{11}\right)$$

$$= \frac{27}{11}$$

$$\alpha_1 = \frac{20}{121}$$

$$\alpha_2 = \frac{26}{121}$$

$$\alpha_3 = -\frac{6}{121}$$

$$\vec{w} = \left(\frac{2}{11}, \frac{6}{11}\right)$$

$$N = 3$$

$$\vec{x}_1 = (2,2)$$

$$\vec{x}_2 = (4,5)$$

$$\vec{x}_3 = (7,4)$$

$$y_1 = -1$$

$$y_2 = +1$$

$$y_3 = +1$$

$$-\alpha_1 + \alpha_2 + \alpha_3 = 0$$

SVM Classifier to Find Hyperplane – Solved Example – Step 4

- The SVM classifier function is given by

$$f(\vec{x}) = \vec{w} \cdot \vec{x} - b$$

- Where,

- $\vec{x} = (x_1, x_2)$

$$= \frac{2}{11}x_1 + \frac{6}{11}x_2 - \frac{27}{11}$$

- The equation of the maximal margin hyperplane is

$$f(\vec{x}) = 0 \qquad f(\vec{x}) = \frac{2}{11}x_1 + \frac{6}{11}x_2 - \frac{27}{11}$$

$$\alpha_1 = \frac{20}{121}$$

$$\alpha_2 = \frac{26}{121}$$

$$\alpha_3 = -\frac{6}{121}$$

$$\vec{w} = \left(\frac{2}{11}, \frac{6}{11}\right)$$

$$b = \frac{27}{11}$$

$$N = 3$$
$$\vec{x}_1 = (2, 2)$$
$$\vec{x}_2 = (4, 5)$$
$$\vec{x}_3 = (7, 4)$$
$$y_1 = -1$$
$$y_2 = +1$$
$$y_3 = +1$$
$$-\alpha_1 + \alpha_2 + \alpha_3 = 0$$

SVM Classifier to Find Hyperplane – Solved Example – Step 5

$$x_1 + 3x_2 - \frac{27}{2} = 0$$