

Support Vector Machines

Nipun Batra

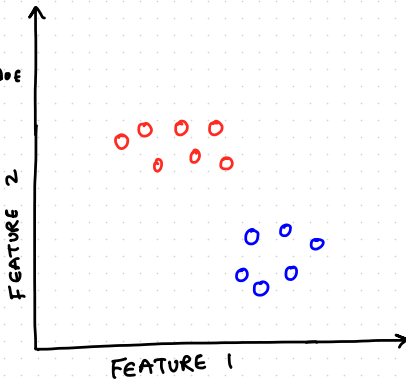
June 22, 2020

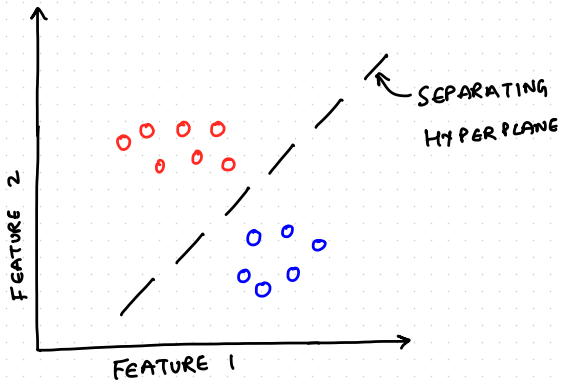
IIT Gandhinagar

SUPPORT VECTOR MACHINES

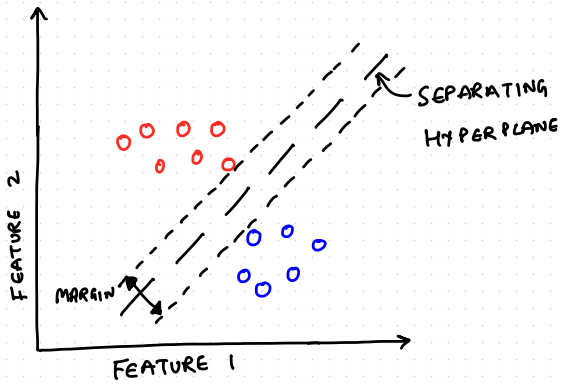
POPULAR BINARY

CLASSIFICATION TECHNIQUE

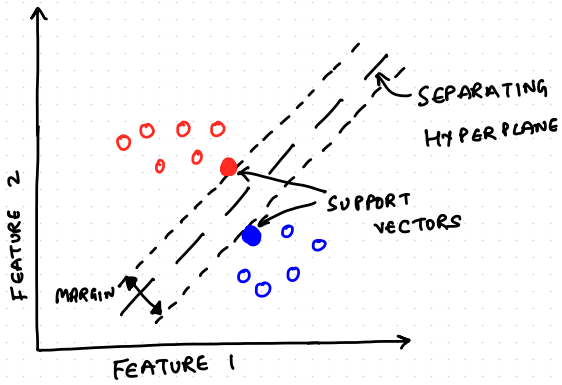




IDEA: DRAW A SEPARATING HYPER PLANE



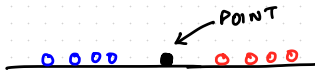
IDEA: MAXIMIZE THE MARGIN



SUPPORT VECTORS: POINTS ON BOUNDARY | MARGIN

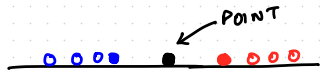
HYPERPLANE VS # DIMENSIONS

1D

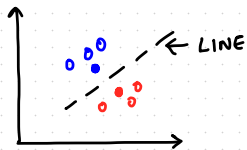


HYPERPLANE VIS # DIMENSIONS

1D

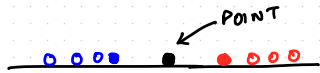


2D

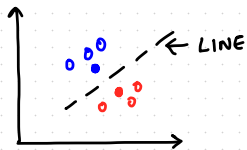


HYPERPLANE VIS # DIMENSIONS

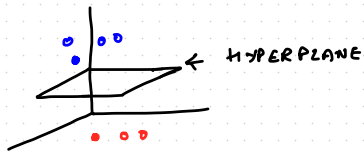
1D



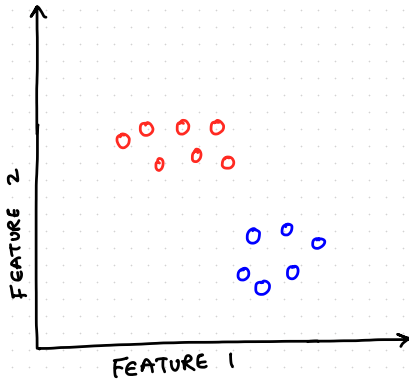
2D



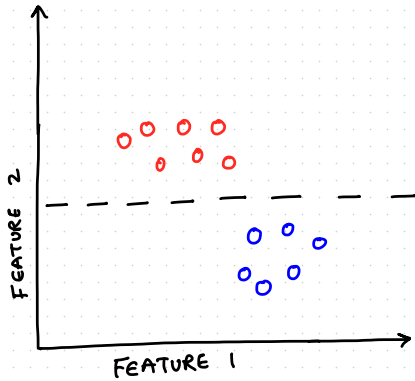
3D
(AND MORE)



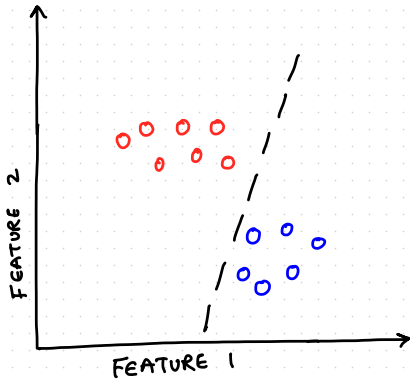
WHICH HYPERPLANE?



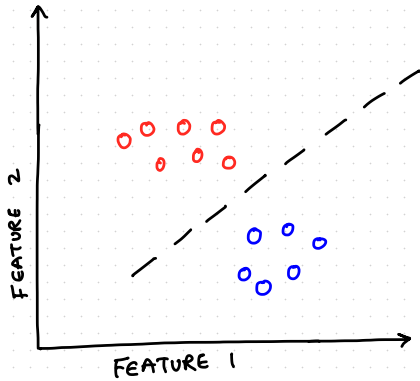
WHICH HYPERPLANE?



WHICH HYPERPLANE?

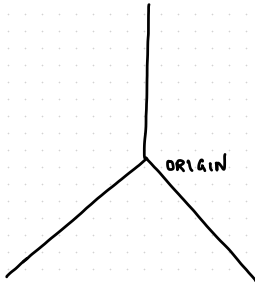


WHICH HYPERPLANE?

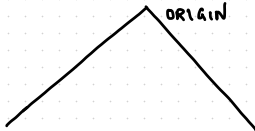


EQUATION OF HYPERPLANE

How TO DEFINE?

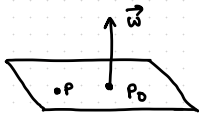


EQUATION OF HYPERPLANE

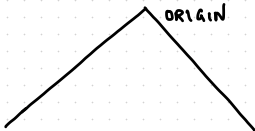


P : Any point on plane
 P_0 : One point on plane

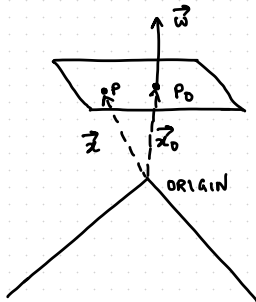
EQUATION OF HYPERPLANE



\vec{w} : \perp vector to
plane at P_0

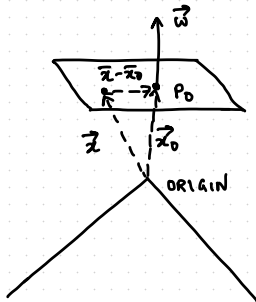


EQUATION OF HYPERPLANE



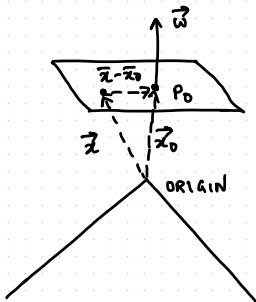
P and P_0 lie on
plane

EQUATION OF HYPERPLANE



$\vec{w} \cdot (\vec{x} - \vec{x}_0) = 0$ lies on plane

EQUATION OF HYPERPLANE



$\vec{P} P_0 = \vec{x} - \vec{x}_0$ lies on
plane

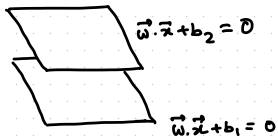
$$\Rightarrow \vec{w} \perp (\vec{x} - \vec{x}_0)$$

$$\text{or, } \vec{w} \cdot (\vec{x} - \vec{x}_0) = 0$$

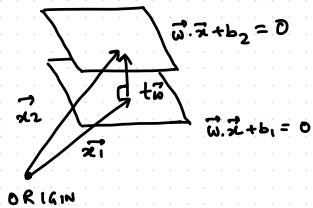
$$\text{or, } \vec{w} \cdot \vec{x} - \vec{w} \cdot \vec{x}_0 = 0$$

$$\text{or, } \boxed{\vec{w} \cdot \vec{x} + b = 0}$$

DISTANCE B/W || HYPER PLANES



DISTANCE B/W || HYPER PLANES



Distance between 2 parallel hyperplanes

Equation of two planes is:

$$\vec{w} \cdot \vec{x} + b_1 = 0$$

$$\vec{w} \cdot \vec{x} + b_2 = 0$$

Distance between 2 parallel hyperplanes

Equation of two planes is:

$$\vec{w} \cdot \vec{x} + b_1 = 0$$

$$\vec{w} \cdot \vec{x} + b_2 = 0$$

For a point \vec{x}_1 on plane 1 and \vec{x}_2 on plane 2, we have:

Distance between 2 parallel hyperplanes

Equation of two planes is:

$$\vec{w} \cdot \vec{x} + b_1 = 0$$

$$\vec{w} \cdot \vec{x} + b_2 = 0$$

For a point \vec{x}_1 on plane 1 and \vec{x}_2 on plane 2, we have:

$$\vec{x}_2 = \vec{x}_1 + t\vec{w}$$

$$D = |t\vec{w}| = |t| \|\vec{w}\|$$

Distance between 2 parallel hyperplanes

Equation of two planes is:

$$\vec{w} \cdot \vec{x} + b_1 = 0$$

$$\vec{w} \cdot \vec{x} + b_2 = 0$$

For a point \vec{x}_1 on plane 1 and \vec{x}_2 on plane 2, we have:

$$\vec{x}_2 = \vec{x}_1 + t\vec{w}$$

$$D = |t\vec{w}| = |t| \|\vec{w}\|$$

We can rewrite as follows:

Distance between 2 parallel hyperplanes

Equation of two planes is:

$$\vec{w} \cdot \vec{x} + b_1 = 0$$

$$\vec{w} \cdot \vec{x} + b_2 = 0$$

For a point \vec{x}_1 on plane 1 and \vec{x}_2 on plane 2, we have:

$$\vec{x}_2 = \vec{x}_1 + t\vec{w}$$

$$D = |t\vec{w}| = |t| \|\vec{w}\|$$

We can rewrite as follows:

$$\vec{w} \cdot \vec{x}_2 + b_2 = 0$$

$$\Rightarrow \vec{w} \cdot (\vec{x}_1 + t\vec{w}) + b_2 = 0$$

Distance between 2 parallel hyperplanes

Equation of two planes is:

$$\vec{w} \cdot \vec{x} + b_1 = 0$$

$$\vec{w} \cdot \vec{x} + b_2 = 0$$

For a point \vec{x}_1 on plane 1 and \vec{x}_2 on plane 2, we have:

$$\vec{x}_2 = \vec{x}_1 + t\vec{w}$$

$$D = |t\vec{w}| = |t|\|\vec{w}\|$$

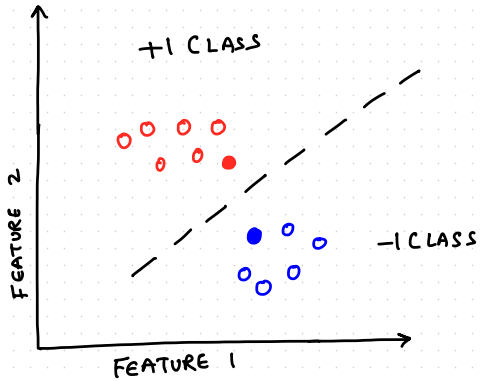
We can rewrite as follows:

$$\vec{w} \cdot \vec{x}_2 + b_2 = 0$$

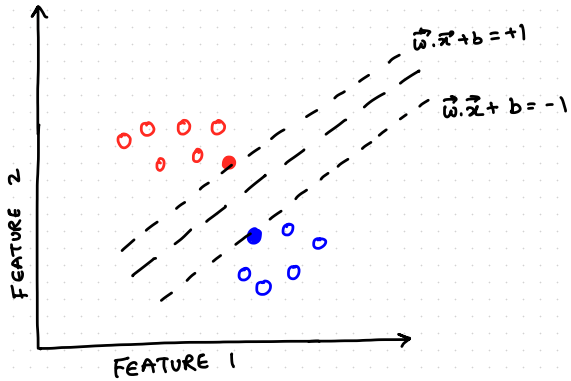
$$\Rightarrow \vec{w} \cdot (\vec{x}_1 + t\vec{w}) + b_2 = 0$$

$$\Rightarrow \vec{w} \cdot \vec{x}_1 + t\|\vec{w}\|^2 + b_1 - b_1 + b_2 = 0 \Rightarrow t = \frac{b_1 - b_2}{\|\vec{w}\|^2} \Rightarrow D = t\|\vec{w}\| = \frac{b_1 - b_2}{\|\vec{w}\|}$$

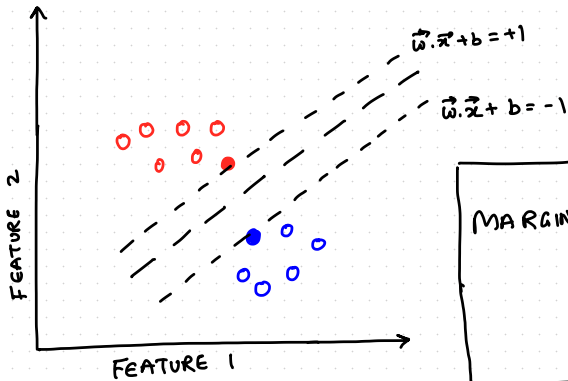
FORMULATION



FORMULATION

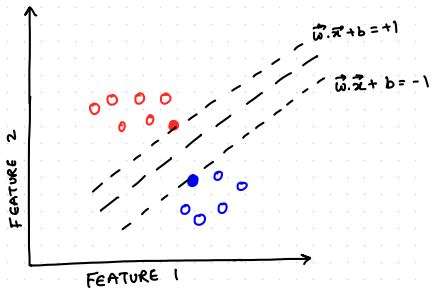


FORMULATION



$$\begin{aligned} \text{MARGIN} &= \frac{(b+1) - (b-1)}{\|\vec{w}\|} \\ &= \frac{2}{\|\vec{w}\|} \end{aligned}$$

FORMULATION



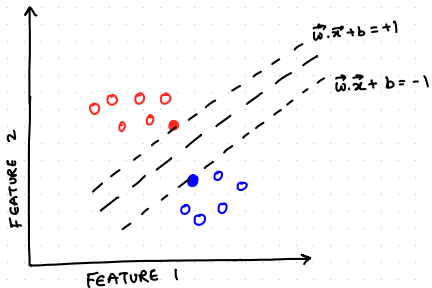
GOAL: MAXIMIZE MARGIN

$$\Rightarrow \text{MAXIMIZE } \frac{2}{\|\vec{w}\|}$$

$$\Rightarrow \text{MINIMIZE } \|\vec{w}\|$$

S.T. Correctly label points

FORMULATION



GOAL: MAXIMIZE MARGIN

$$\Rightarrow \text{MAXIMIZE } \frac{2}{\|\vec{w}\|}$$

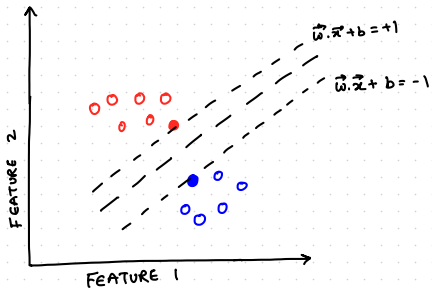
$$\Rightarrow \text{MINIMIZE } \|\vec{w}\|$$

S.T. Correctly label points

i.e. if $y_i = -1$
 $\vec{w} \cdot \vec{x} + b \leq -1$

if $y_i = +1$
 $\vec{w} \cdot \vec{x} + b \geq +1$

FORMULATION



GOAL: MAXIMIZE MARGIN

$$\Rightarrow \text{MAXIMIZE } \frac{2}{\|\vec{w}\|}$$

$$\Rightarrow \boxed{\text{MINIMIZE } \|\vec{w}\|}$$

S.T. Correctly label points

i.e. if $y_i = -1$
 $\vec{w} \cdot \vec{x} + b \leq -1$

if $y_i = +1$
 $\vec{w} \cdot \vec{x} + b \geq 1$

$$\boxed{y_i (\vec{w} \cdot \vec{x} + b) \geq 1}$$

Primal Formulation

Objective

$$\begin{aligned} \text{Minimize } & \frac{1}{2} \|w\|^2 \\ \text{s.t. } & y_i(w \cdot x_i + b) \geq 1 \quad \forall i \end{aligned}$$

Primal Formulation

Objective

$$\begin{aligned} \text{Minimize } & \frac{1}{2} \|w\|^2 \\ \text{s.t. } & y_i(w \cdot x_i + b) \geq 1 \quad \forall i \end{aligned}$$

Q) What is $\|w\|$?

Primal Formulation

Objective

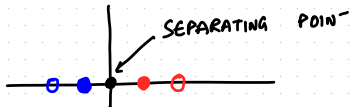
$$\begin{aligned} \text{Minimize } & \frac{1}{2} \|w\|^2 \\ \text{s.t. } & y_i(w \cdot x_i + b) \geq 1 \quad \forall i \end{aligned}$$

Q) What is $\|w\|$?

$$w = \begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_n \end{bmatrix}$$

$$\begin{aligned} \|w\| &= \sqrt{w^T w} \\ &= \sqrt{\begin{bmatrix} w_1, w_2, \dots, w_n \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_n \end{bmatrix}} \end{aligned}$$

EXAMPLE (IN 1D)



Simple Exercise

$$\begin{bmatrix} x & y \\ 1 & 1 \\ 2 & 1 \\ -1 & -1 \\ -2 & -1 \end{bmatrix}$$

Separating Hyperplane: $wx + b = 0$

Simple Exercise

$$y_i(w_i x_i + b) \geq 1$$

$$\begin{bmatrix} x_1 & y \\ 1 & 1 \\ 2 & 1 \\ -1 & -1 \\ -2 & -1 \end{bmatrix}$$

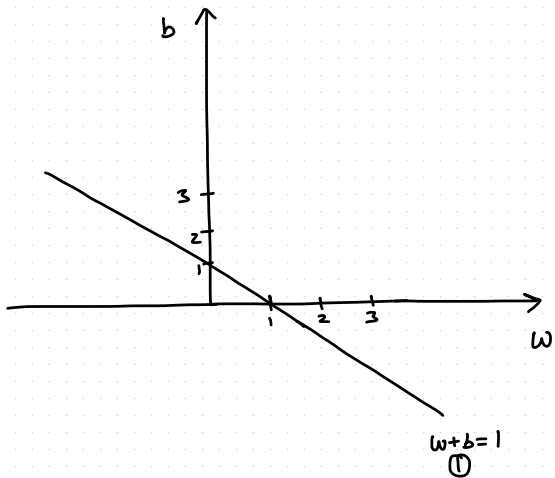
$$\Rightarrow y_i(w_i x_i + b) \geq 1$$

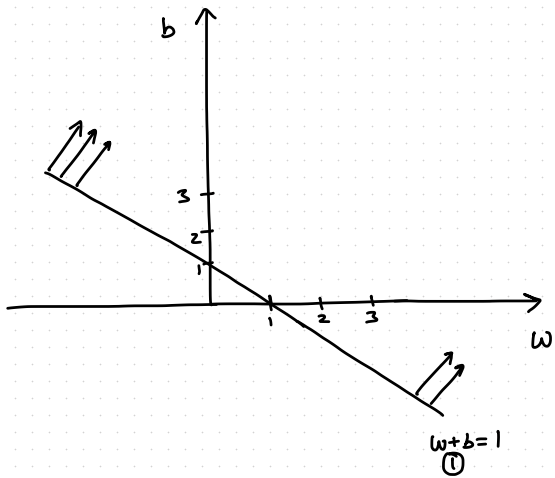
$$\Rightarrow 1(w_1 + b) \geq 1$$

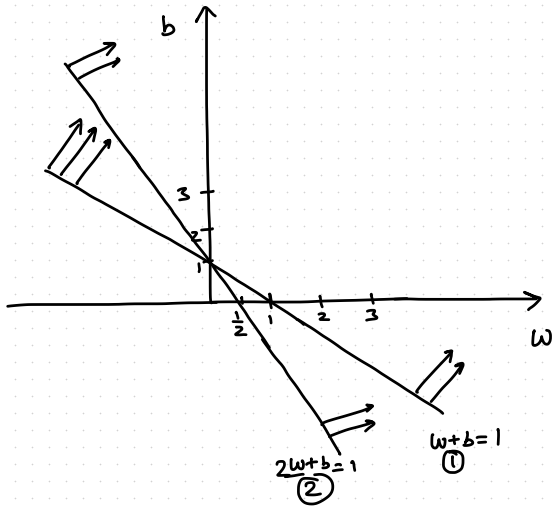
$$\Rightarrow 1(2w_1 + b) \geq 1$$

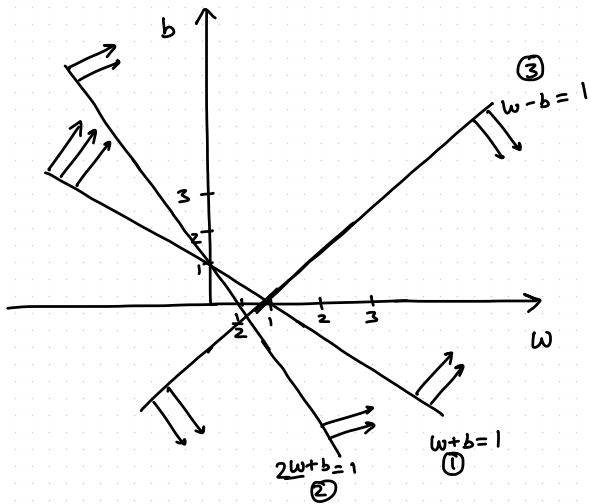
$$\Rightarrow -1(-w_1 + b) \geq 1$$

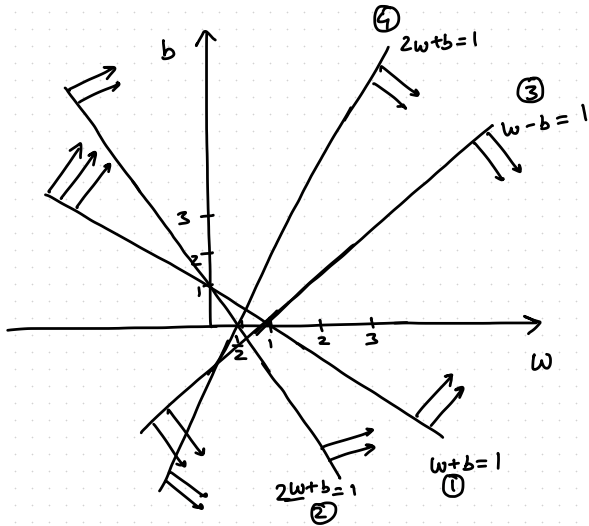
$$\Rightarrow -1(-2w_1 + b) \geq 1$$

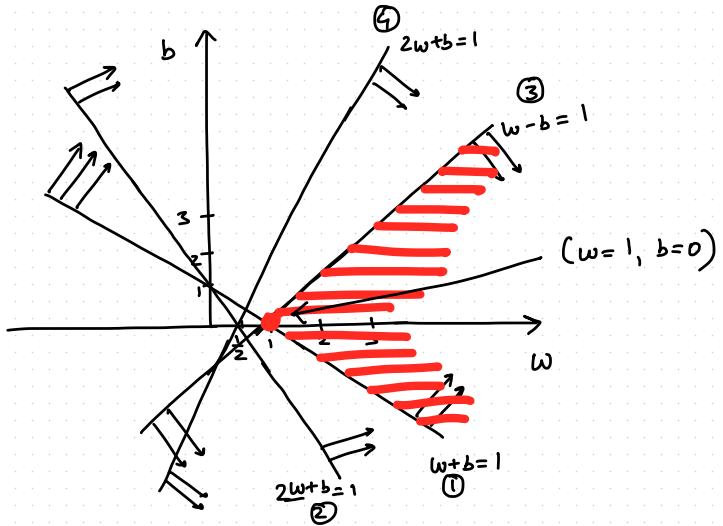












Simple Exercise

$$w_{min} = 1, b = 0$$

$$w \cdot x + b = 0$$

$$x = 0$$

Simple Exercise

Minimum values satisfying constraints $\Rightarrow w = 1$ and $b = 0$

\therefore Max margin classifier $\Rightarrow x = 0$

Primal Formulation is a Quadratic Program

Generally;

⇒ Minimize Quadratic(x)

⇒ such that, Linear(x)

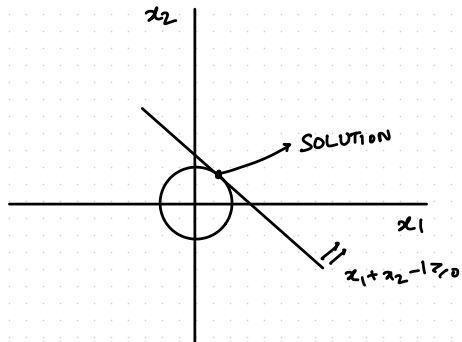
Question

$$x = (x_1, x_2)$$

$$\text{minimize } \frac{1}{2} \|x\|^2$$

$$: x_1 + x_2 - 1 \geq 0$$

MINIMIZE QUADRATIC
S.t. LINEAR



Converting to Dual Problem

Primal \Rightarrow Dual Conversion using Lagrangian multipliers

$$\begin{aligned} &\text{Minimize } \frac{1}{2} \|\bar{w}\|^2 \\ &\text{s.t. } y_i(\bar{w} \cdot x_i + b) \geq 1 \\ &\quad \forall i \end{aligned}$$

$$L(\bar{w}, b, \alpha_1, \alpha_2, \dots, \alpha_n) = \frac{1}{2} \sum_{i=1}^d w_i^2 - \sum_{i=1}^N \alpha_i (y_i(\bar{w} \cdot \bar{x}_i + b) - 1) \quad \forall \alpha_i \geq 0$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^n \alpha_i y_i = 0$$

Converting to Dual Problem

$$\frac{\partial L}{\partial w} = 0 \Rightarrow \bar{w} - \sum_{i=1}^n \alpha_i y_i \bar{x}_i = 0$$

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

$$\begin{aligned} L(\bar{w}, b, \alpha_1, \alpha_2, \dots, \alpha_n) &= \frac{1}{2} \sum_{i=1}^d w_i^2 - \sum_{i=1}^N \alpha_i (y_i (\bar{w} \cdot \bar{x}_i + b) - 1) \\ &= \frac{1}{2} \|\bar{w}\|^2 - \sum_{i=1}^N \alpha_i y_i \bar{w} \cdot \bar{x}_i - \sum_{i=1}^N \alpha_i y_i b + \sum_{i=1}^N \alpha_i \\ &= \sum_{i=1}^N \alpha_i + \frac{(\sum_i \alpha_i y_i \bar{x}_i) (\sum_j \alpha_j y_j \bar{x}_j)}{2} - \sum_i \alpha_i y_i \left(\sum_j \alpha_j y_j \bar{x}_j \right) \bar{x}_i \end{aligned}$$

Converting to Dual Problem

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

Minimize $\|\bar{w}\|^2 \Rightarrow$ Maximize $L(\alpha)$

s.t

$$y_i (\bar{w}, x_i + b) \geq 1$$

s.t

$$\sum_{i=1}^N \alpha_i y_i = 0 \quad \forall \alpha_i \geq 0$$

Question

Question:

$$\alpha_i (y_i (\bar{w} \cdot \bar{x}_i + b) - 1) = 0 \quad \forall i \text{ as per KKT slackness}$$

What is α_i for support vector points?

Answer: For support vectors,

$$\bar{w} \cdot \bar{x}_i + b = -1 \text{ (+ve class)}$$

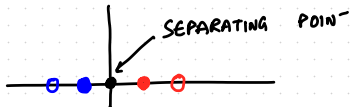
$$\bar{w} \cdot \bar{x}_i + b = +1 \text{ (+ve class)}$$

$$y_i (\bar{w} \cdot \bar{x}_i + b) - 1 = 0 \quad \text{for } i = \{\text{support vector points}\}$$

$$\therefore \alpha_i \text{ where } i \in \{\text{support vector points}\} \neq 0$$

For all non-support vector points $\alpha_i = 0$

EXAMPLE (IN 1D)



Revisiting the Simple Example

$$\begin{bmatrix} x_1 & y \\ 1 & 1 \\ 2 & 1 \\ -1 & -1 \\ -2 & -1 \end{bmatrix}$$

$$L(\alpha) = \sum_{i=1}^4 \alpha_i - \frac{1}{2} \sum_{i=1}^4 \sum_{j=1}^4 \alpha_i \alpha_j y_i y_j \bar{x}_i \bar{x}_j \quad \alpha_i \geq 0$$

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

Revisiting the Simple Example

$$\begin{aligned}L(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = & \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 \\ & - \frac{1}{2} \{ \alpha_1 \alpha_1 \times (1 * 1) \times (1 * 1) \\ & + \\ & \alpha_1 \alpha_2 \times (1 * 1) \times (1 * 2) \\ & + \\ & \alpha_1 \alpha_3 \times (1 * -1) \times (1 * 1) \\ & \dots \\ & \alpha_4 \alpha_4 \times (-1 * -1) \times (-2 * -2) \}\end{aligned}$$

How to Solve? \Rightarrow Use the QP Solver!!

Revisiting the Simple Example

For the trivial example,

We know that only $x = \pm 1$ will take part in the constraint actively. Thus, $\alpha_2, \alpha_4 = 0$

By symmetry, $\alpha_1 = \alpha_3 = \alpha$ (say)

$$\& \sum y_i \alpha_i = 0$$

$$L(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = 2\alpha$$

$$\begin{aligned} & - \frac{1}{2} \{ \alpha^2(1)(-1)(1)(-1) \\ & \quad + \alpha^2(-1)(1)(-1)(1) \\ & \quad + \alpha^2(1)(1)(1)(1) + \alpha^2(-1)(-1)(-1)(-1) \\ & \quad \} \end{aligned}$$

$$\underset{\alpha}{\text{Maximize}} \quad 2\alpha - \frac{1}{2}(4\alpha^2)$$

Revisiting the Simple Example

$$\begin{aligned}\frac{\partial}{\partial \alpha} (2\alpha - 2\alpha^2) &= 0 \Rightarrow 2 - 4\alpha = 0 \\ &\Rightarrow \alpha = 1/2\end{aligned}$$

$$\therefore \alpha_1 = 1/2 \quad \alpha_2 = 0; \quad \alpha_3 = 1/2 \quad \alpha_4 = 0$$

$$\begin{aligned}\vec{w} &= \sum_{i=1}^N \alpha_i y_i \bar{x}_i = 1/2 \times 1 \times 1 + 0 \times 1 \times 2 \\ &\quad + 1/2 \times -1 \times -1 + 0 \times -1 \times -2 \\ &= 1/2 + 1/2 = 1\end{aligned}$$

Revisiting the Simple Example

Finding b:

For the support vectors we have,

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

$$\text{or, } y_i (\bar{w} \cdot \bar{x}_1 + b) = 1$$

$$\text{or, } y_i^2 (\bar{w} \cdot \bar{x}_i + b) = y_i$$

$$\text{or, } \bar{w} \cdot \bar{x}_i + b = y_i \quad (\because y_i^2 = 1)$$

$$\text{or, } b = y_i - w \cdot x_i$$

$$\text{In practice, } b = \frac{1}{N_{SV}} \sum_{i=1}^{N_{SV}} (y_i - \bar{w} \bar{x}_i)$$

Obtaining the Solution

$$\begin{aligned} b &= \frac{1}{2}\{(1 - (1)(1)) + (-1 - (1)(-1))\} \\ &= \frac{1}{2}\{0 + 0\} = 0 \\ &= 0 \\ \therefore w &= 1 \text{ \& } b = 0 \end{aligned}$$

Making Predictions

$$\hat{y}(x_i) = \text{SIGN}(w \cdot x_i + b)$$

For $x_{test} = 3$; $\hat{y}(3) = \text{SIGN}(1 \times 3 + 0) = +ve$ class

Making Predictions

Alternatively,

$$\begin{aligned}\hat{y}(x_{TEST}) &= \text{SIGN}(\bar{w} \cdot \bar{x}_{TEST} + b) \\ &= \text{SIGN}\left(\sum_{i=1}^{N_S} \alpha_j y_j x_j \cdot x_{test} + b\right)\end{aligned}$$

In our example,

$$\alpha_1 = 1/2; \alpha_2 = 0; \quad \alpha_3 = 1/2; \alpha_4 = 0$$

$$\begin{aligned}\hat{y}(3) &= \text{SIGN}\left(\frac{1}{2} \times 1 \times (1 \times 3) + 0 + \frac{1}{2} \times (-1) \times (-1 \times 3) + 0\right) \\ &= \text{SIGN}\left(\frac{6}{2}\right) = \text{SIGN}(3) = +1\end{aligned}$$