Support Vector Machines

Nipun Batra

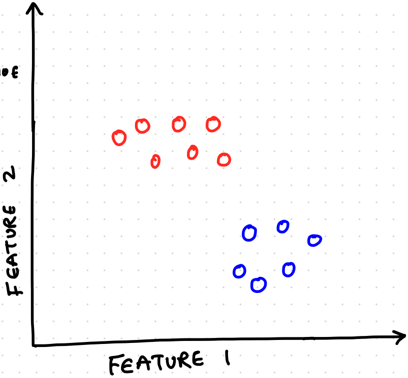
June 16, 2020

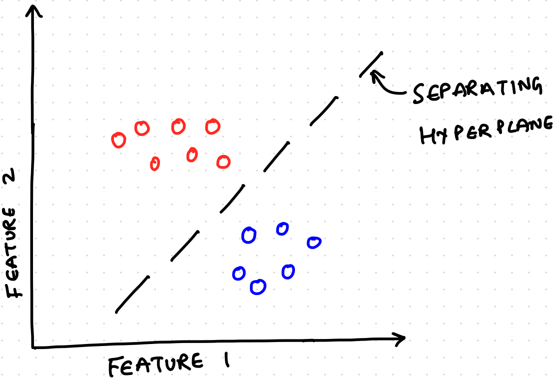
IIT Gandhinagar

SUPPORT VECTOR MACHINES

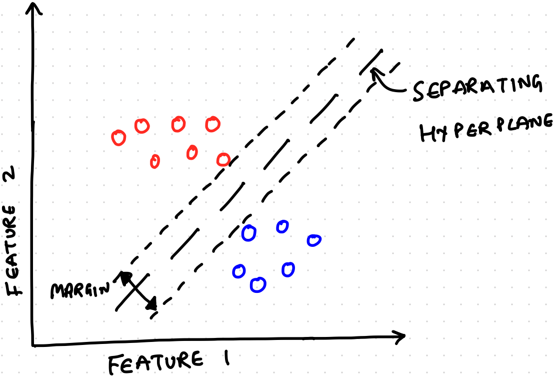
POPULAR BINARY

CLASSIFICATION TECHNIQUE

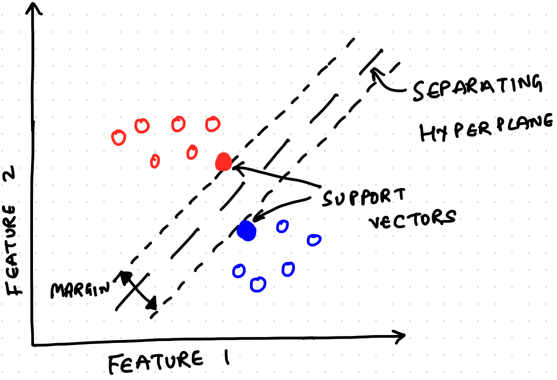




IDEA: DRAW A SEPARATING HYPERPLANE



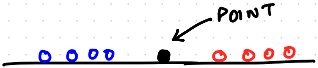
IDEA: MAXIMIZE THE MARGIN



SUPPORT VECTORS: POINTS ON BOUNDARY | MARGIN

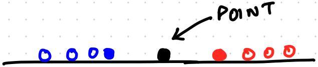
HYPERPLANE V/S # DIMENSIONS

1D

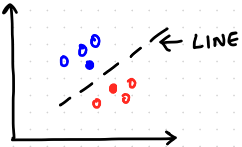


HYPERPLANE V/S # DIMENSIONS

1D

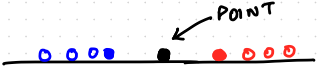


2D

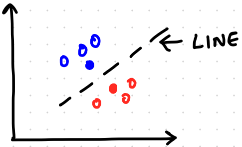


HYPERPLANE V/S # DIMENSIONS

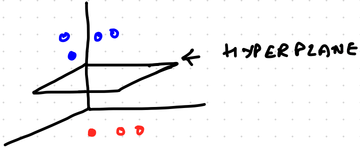
1D



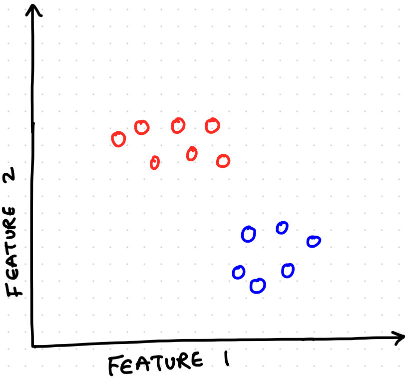
2D



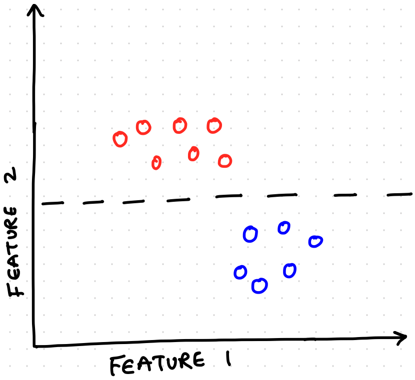
3D
(AND
MORE)



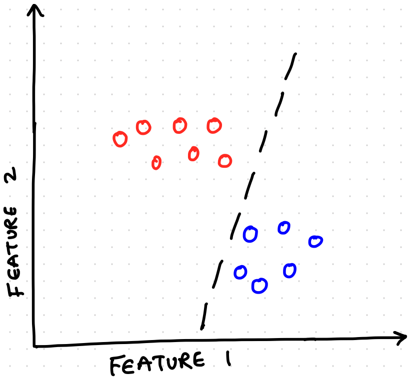
WHICH HYPERPLANE?



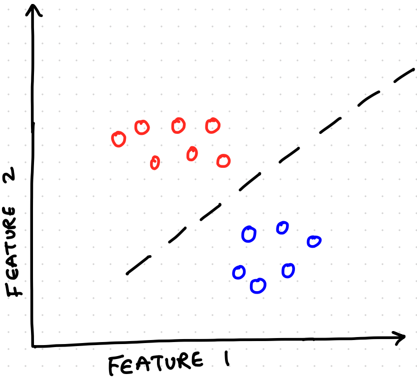
WHICH HYPERPLANE?



WHICH HYPERPLANE?

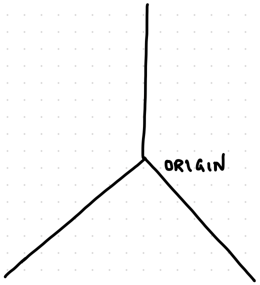


WHICH HYPERPLANE?

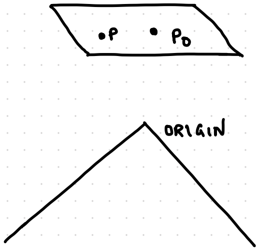


EQUATION OF HYPERPLANE

HOW TO DEFINE?



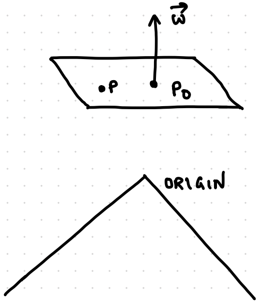
EQUATION OF HYPERPLANE



P : Any point on plane

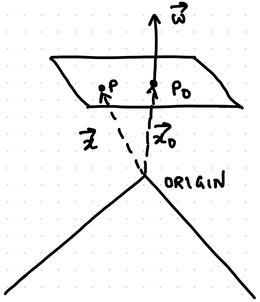
P_0 : One point on plane

EQUATION OF HYPERPLANE



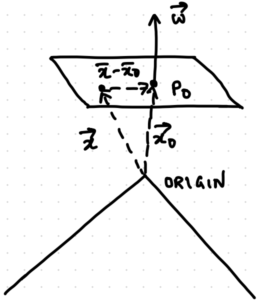
$\vec{w}$: $\perp$ vector to plane at P_0

EQUATION OF HYPERPLANE



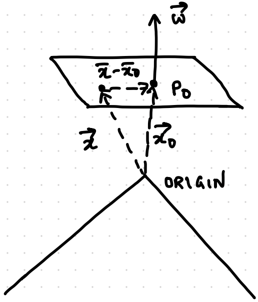
P and P_0 lie on plane

EQUATION OF HYPERPLANE



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

EQUATION OF HYPERPLANE



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

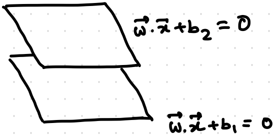
$\Rightarrow \vec{w} \perp (\vec{x} - \vec{x}_0)$

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

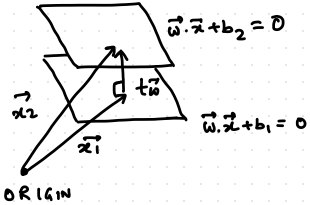
or, $\vec{w} \cdot \vec{x} - \vec{w} \cdot \vec{x}_0 = 0$

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

DISTANCE B/W ∥ HYPERPLANES



DISTANCE B/W || HYPERPLANES



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:

$$\overrightarrow{x_2} = \overrightarrow{x_1} + t\vec{w}$$
$$D = |t\vec{w}| = |t|||\vec{w}||$$

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:

$$\overrightarrow{x_2} = \overrightarrow{x_1} + t\vec{w}$$

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

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:

$$\overrightarrow{x_2} = \overrightarrow{x_1} + t\vec{w}$$
$$D = |t\vec{w}| = |t|||\vec{w}||$$

We can rewrite as follows:

$$\vec{w} \cdot \vec{x}_2 + b_2 = 0$$
$$\Rightarrow \vec{w} \cdot \left(\vec{x}_1 + t\vec{w}\right) + b_2 = 0$$

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:

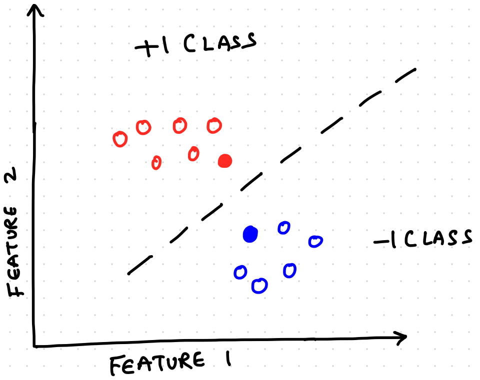
$$\overrightarrow{x_2} = \overrightarrow{x_1} + t\vec{w}$$
$$D = |t\vec{w}| = |t|||\vec{w}||$$

We can rewrite as follows:

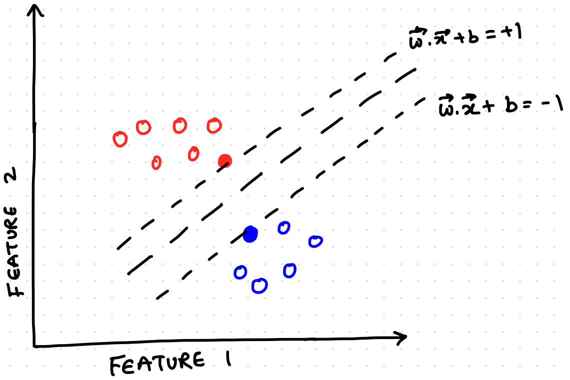
$$\vec{w} \cdot \vec{x}_2 + b_2 = 0$$
$$\Rightarrow \vec{w} \cdot \left(\vec{x}_1 + t\vec{w}\right) + b_2 = 0$$

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

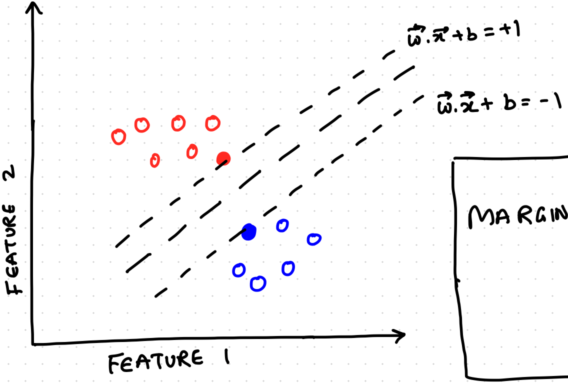
FORMULATION



FORMULATION



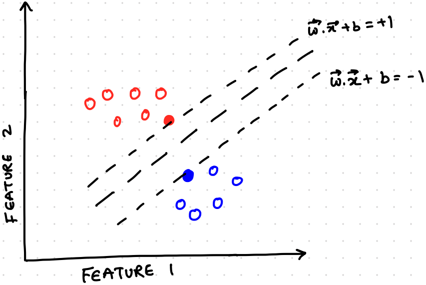
FORMULATION



$$\text{MARGIN} = \frac{(b+1)-(b-1)}{\|\vec{w}\|}$$

$$= \frac{2}{\|\vec{w}\|}$$

FORMULATION



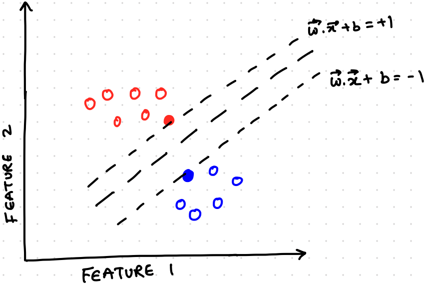
GOAL: MAXIMIZE MARGIN

$\Rightarrow$ MAXIMIZE $\dfrac{2}{||\vec{w}||}$

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

S.T. Correctly label points

FORMULATION



GOAL: MAXIMIZE MARGIN

$\Rightarrow$ MAXIMIZE $\dfrac{2}{\|\vec{w}\|}$

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

S.T. Correctly label points

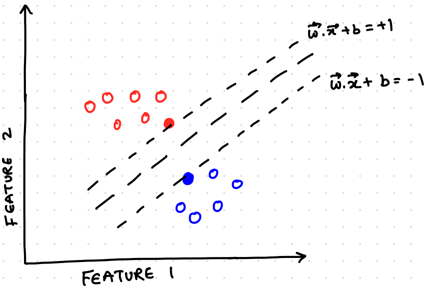
i.e. if $y_i = -1$

$\vec{w}.\vec{x}+b \leq -1$

if $y_i = +1$

$\vec{w}.\vec{x}+b \geq 1$

FORMULATION



GOAL: MAXIMIZE MARGIN

$\Rightarrow$ MAXIMIZE $\dfrac{2}{||\vec{w}||}$

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

S.T. correctly label points

i.e. if $y_i = -1$

$\vec{w}.\vec{x}+b \leq -1$

if $y_i = +1$

$\vec{w}.\vec{x}+b \geq 1$

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

Primal Formulation

Objective

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

Primal Formulation

Objective

$$\text{Minimize } \frac{1}{2}||w||^2$$
$$\text{s.t. } y_i(w.x_i + b) \geq 1 \;\; \forall i$$

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

Primal Formulation

Objective

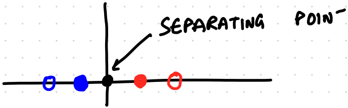
$$\text{Minimize } \frac{1}{2}||w||^2$$
$$\text{s.t. } y_i(w.x_i + b) \geq 1 \ \forall i$$

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

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

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

EXAMPLE (IN 1D)



Simple Exercise

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

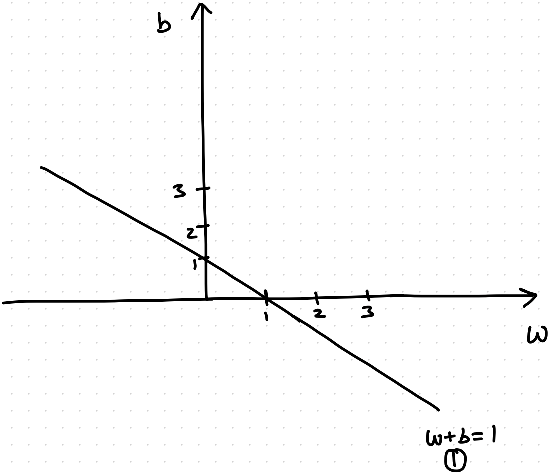
Separating Hyperplane: $wx + b = 0$

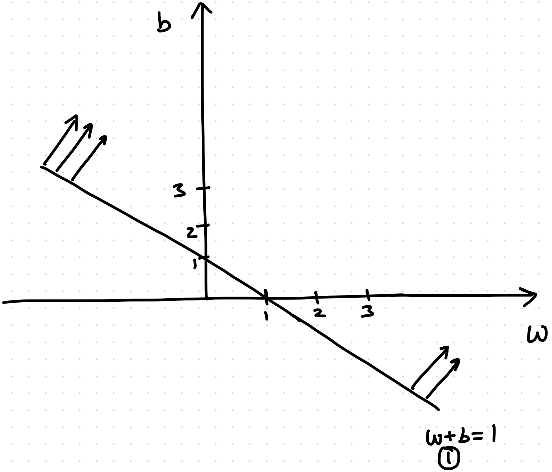
Simple Exercise

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

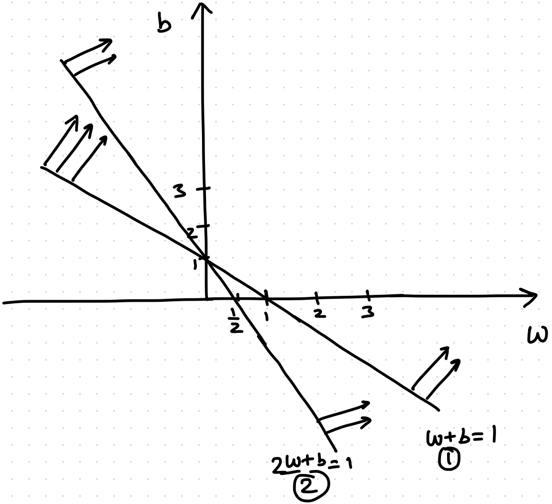
$$\begin{bmatrix} x_1 & y \\ 1 & 1 \\ 2 & 1 \\ -1 & -1 \\ -2 & -1 \end{bmatrix}$$

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

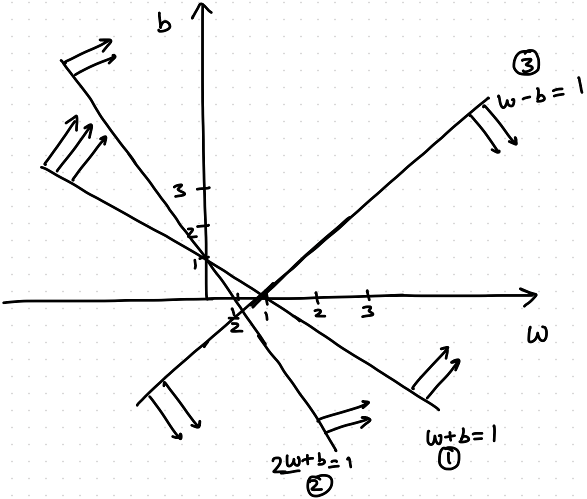

b
3
2
1
1
2
3
w
w+b=1
1

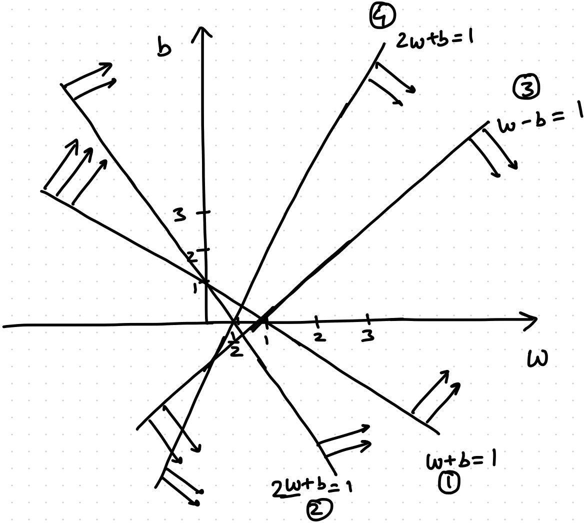


b
3
2
1
1
2
3
w
w+b=1
1

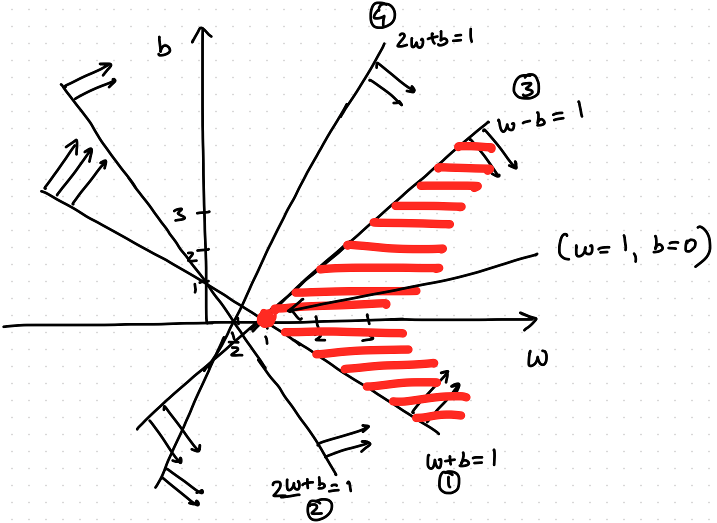


b
3
2
1
1/2
1
2
3
w
w+b=1
1
2w+b=1
2


b
w
3
2
1
1/2
1
2
3
③
w - b = 1
w+b=1
①
2w+b=1
②



b
w
1
2
3
1
2
3
$\frac{1}{2}$
$2w+b=1$
4
$w-b=1$
3
$w+b=1$
1
$2w+b=1$
2



b
w
3
2
1
1
2
1/2
④ 2w+b=1
③ w-b=1
(w=1, b=0)
w+b=1 ①
2w+b=1 ②

Simple Exercise

$$w_{min} = 1, b = 0$$
$$w.x + b = 0$$
$$x = 0$$

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;

$\Rightarrow$ Minimize Quadratic(x)

$\Rightarrow$ such that, Linear(x)

Question

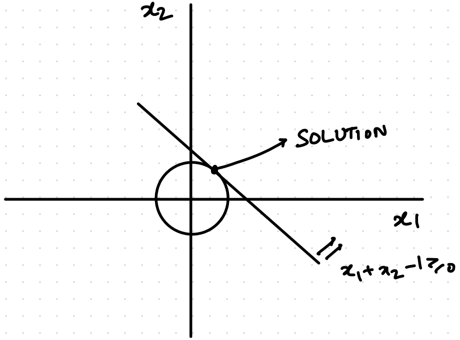
$$x = (x_1, x_2)$$

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

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

MINIMIZE QUADRATIC

S.t. LINEAR



Converting to Dual Problem

Primal $\Rightarrow$ Dual Conversion using Lagrangian multipliers

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

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

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

Converting to Dual Problem

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

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

$$L(\bar{w}, b, \alpha_1, \alpha_2, ...\alpha_n) = \frac{1}{2}\sum_{i=1}^{d} w_i^2 - \sum_{i=1}^{N} \alpha_i (y_i(\bar{w}.\bar{x_i} + b) - 1$$

$$= \frac{1}{2}||\bar{w}||^2 - \sum_{i=1}^{N} \alpha_i y_i \vec{w}.\bar{x_i} - \sum_{i=1}^{N} \alpha_i y_i b + \sum_{i=1}^{N} \alpha_i$$

$$= \sum_{=1}^{N} \alpha_i + \frac{\left(\sum_i \alpha_i y_i \bar{x_i}\right)\left(\sum_j \alpha_j y_j \bar{x_j}\right)}{2} - \sum_i \alpha_i y_i \left(\sum_j \alpha_j y_j \bar{x_j}\right) \bar{x_i}$$

Converting to Dual Problem

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

$$\begin{array}{ll} \text{Minimize } \|\bar{w}\|^2 \Rightarrow & \text{Maximize } L(\alpha) \\ s.t & s.t \\ y_i\left(\bar{w}, x_i + b\right) \geqslant 1 & \sum_{i=1}^{N} \alpha_i y_i = 0 \ \forall \ \alpha_i \geq 0 \end{array}$$

Question:

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

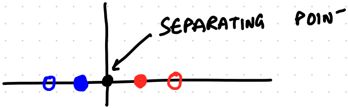
What is α_i for support vector points?

Answer: For support vectors,

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

$y_i \left(\bar{w} \cdot \bar{x}_i + b\right) - 1) = 0 \quad$ for $i = \{$support vector points$\}$
$\therefore \alpha_i$ where i $\in \{$support vector points$\} \neq 0$
For all non-support vector points $\alpha_i = 0$

EXAMPLE (IN 1D)



Revisiting the Simple Example

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

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

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

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 = ± 1 will take part in the constraint actively. Thus, $\alpha_2, \alpha_4 = 0$

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

& $\sum y_i \alpha_i = 0$

$$
\begin{aligned}
L(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = & 2\alpha \\
& - \frac{1}{2} \{ \alpha^2 (1)(-1)(1)(-1) \\
& \quad + \alpha^2 (-1)(1)(-1)(1) \\
& \quad + \alpha^2 (1)(1)(1)(1) + \alpha^2 (-1)(-1)(-1)(-1) \\
& \}
\end{aligned}
$$

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

Revisiting the Simple Example

$$\begin{aligned}
\frac{\partial}{\partial \alpha}\left(2\alpha - 2\alpha^2\right) &= 0 \Rightarrow 2 - 4\alpha = 0 \\
&\Rightarrow \ \alpha = 1/2 \\
\therefore \alpha_1 = 1/2 \ \ \alpha_2 = 0; \ \ \alpha_3 = 1/2 \ \ \alpha_4 &= 0 \\
\vec{w} = \sum_{i=1}^{N} \alpha_i y_i \bar{x}_i &= 1/2 \times 1 \times 1 + 0 \times 1 \times 2 \\
&+1/2 \times -1 \times -1 + 0 \times -1 \times -2 \\
&= 1/2 + 1/2 = 1
\end{aligned}$$

Finding b:

For the support vectors we have,

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

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

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

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

or, $b = y_i - w \cdot x_i$

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

Obtaining the Solution

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

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,

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

In our example,

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

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

$$= \text{SIGN}\left(\frac{6}{2}\right) = \text{SIGN}(3) = +1$$

ORIGINAL DATA
IN R

Non-Linearly Separable Data

- Data not separable in $\mathbb{R}$

Non-Linearly Separable Data

- Data not separable in $\mathbb{R}$
- Can we still use SVM?

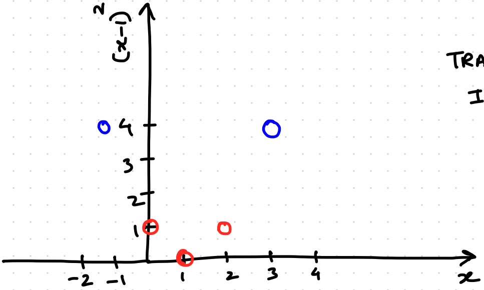
Non-Linearly Separable Data

- Data not separable in $\mathbb{R}$
- Can we still use SVM?
- Yes!

Non-Linearly Separable Data

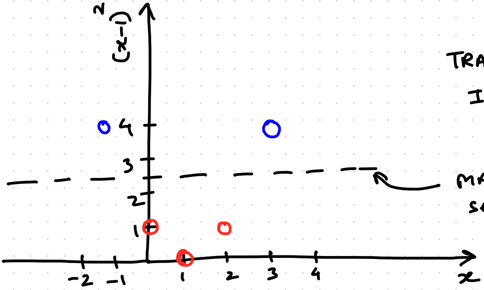
- Data not separable in $\mathbb{R}$
- Can we still use SVM?
- Yes!
- How? Project data to a higher dimensional space.

ORIGINAL DATA
IN R



TRANSFORMED DATA
IN R^2

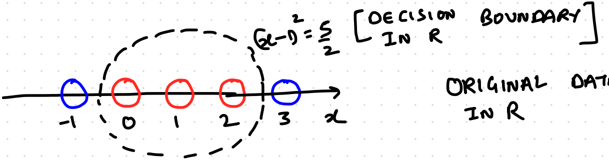
ORIGINAL DATA IN R



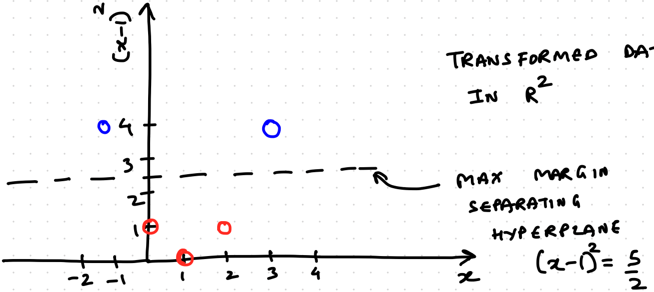
TRANSFORMED DATA IN R^2

MAX MARGIN SEPARATING HYPERPLANE

$$(x-1)^2 = \frac{5}{2}$$

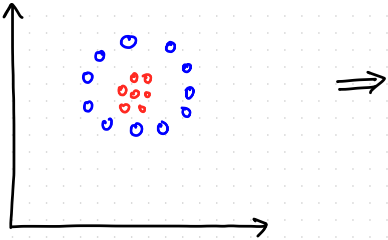


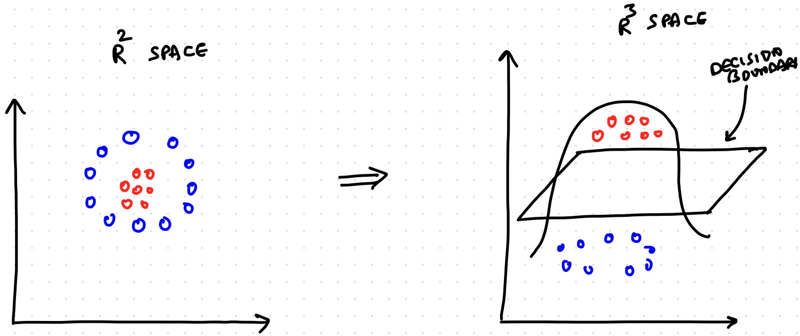
ORIGINAL DATA IN R



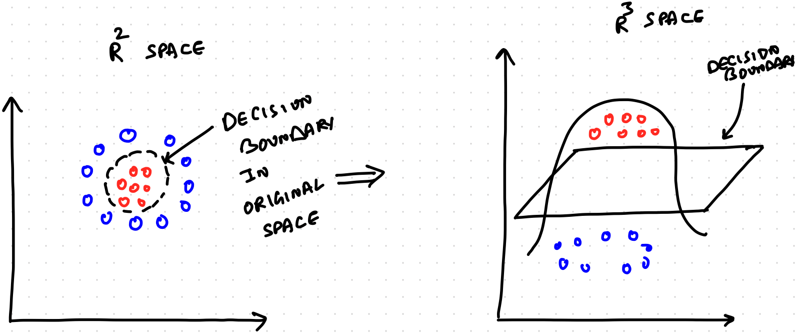
TRANSFORMED DATA IN R^2

R^2 SPACE





R^2 SPACE
R^3 SPACE
DECISION BOUNDARY



R^2 SPACE
DECISION BOUNDARY IN ORIGINAL SPACE
R^3 SPACE
DECISION BOUNDARY

Projection/Transformation Function

$$\phi : \mathbb{R}^d \rightarrow \mathbb{R}^D$$

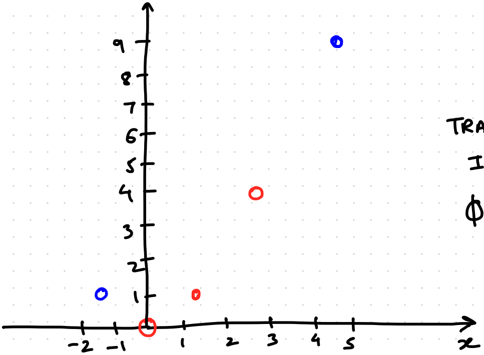
where, d = original dimension

D = new dimension

In example next :

$d = 1; D = 2$

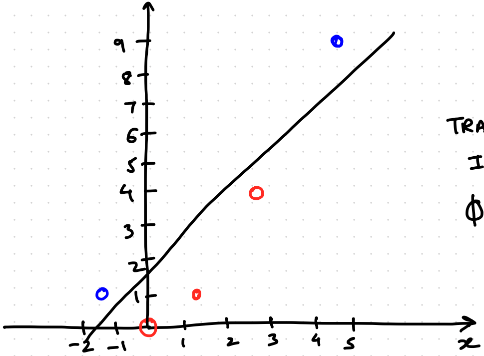
ORIGINAL DATA IN R



TRANSFORMED DATA IN R^2

$$\phi(x) = \begin{bmatrix} \sqrt{2}\,x \\ x^2 \end{bmatrix}$$

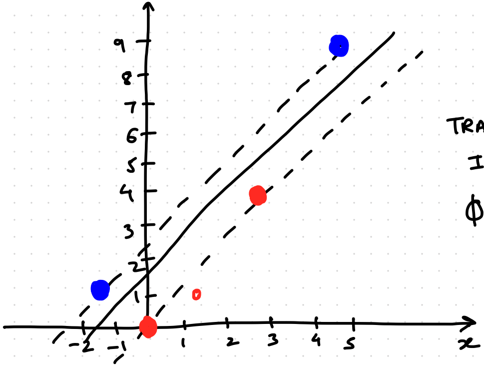
ORIGINAL DATA
IN R



TRANSFORMED DATA
IN R^2

$$\phi(x) = \begin{bmatrix} \sqrt{2}\, x \\ x^2 \end{bmatrix}$$

ORIGINAL DATA
IN R



TRANSFORMED DATA
IN R^2

$$\phi(x) = \begin{bmatrix} \sqrt{2}\,x \\ x^2 \end{bmatrix}$$

Linear SVM:

Maximize

$$L(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \overline{x_i}.\overline{x_j}$$

such that constraints are satisfied.

$$\downarrow$$

Transformation (ϕ)

$$\downarrow$$

$$L(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \phi(\overline{x_i}).\phi(\overline{x_j})$$

Steps

1. Compute $\phi(x)$ for each point

$$\phi : \mathbb{R}^d \to \mathbb{R}^D$$

2. Computer dot products over $\mathbb{R}^D$ space

Steps

1. Compute $\phi(x)$ for each point

$$\phi : \mathbb{R}^d \rightarrow \mathbb{R}^D$$

2. Computer dot products over $\mathbb{R}^D$ space

Q. If $D >> d$

Steps

1. Compute $\phi(x)$ for each point

$$\phi : \mathbb{R}^d \rightarrow \mathbb{R}^D$$

2. Computer dot products over $\mathbb{R}^D$ space

Q. If $D >> d$

Both steps are expensive!

Kernel Trick