SVM-2

z_i $\qquad$ $1-\xi_i$

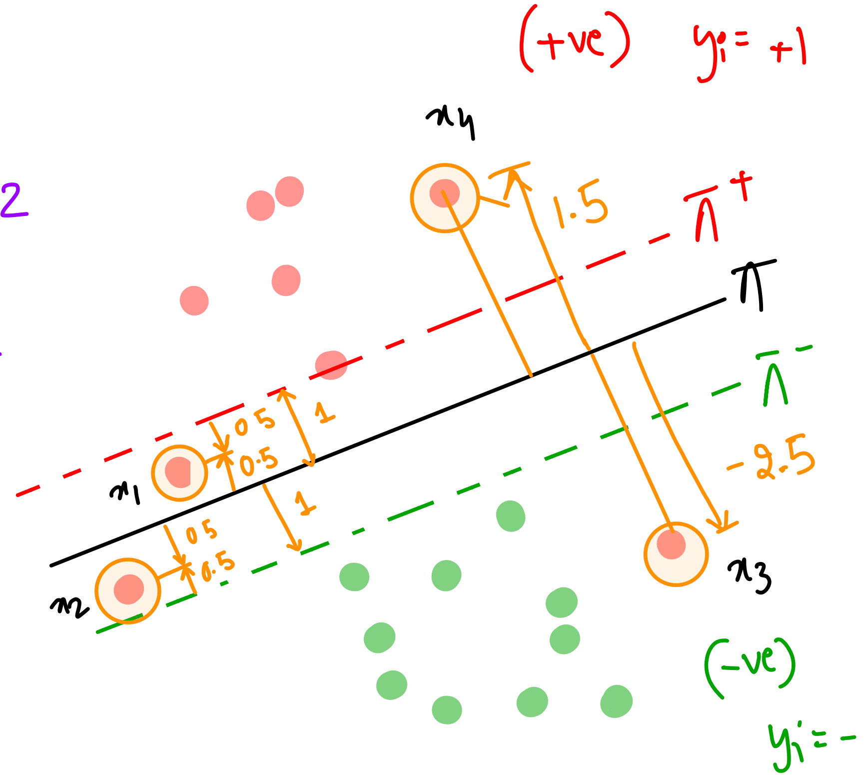
$x_1 \Rightarrow 0.5 \Rightarrow 1-0.5$ $\quad \xi_1$

$x_2 \Rightarrow -0.5 \Rightarrow 1-1.5$ $\quad \xi_2$

$x_3 \Rightarrow -2.5 \Rightarrow 1-3.5$ $\quad \xi_3$

$x_4 \Rightarrow 1.5 \Rightarrow 1-(-0.5)$

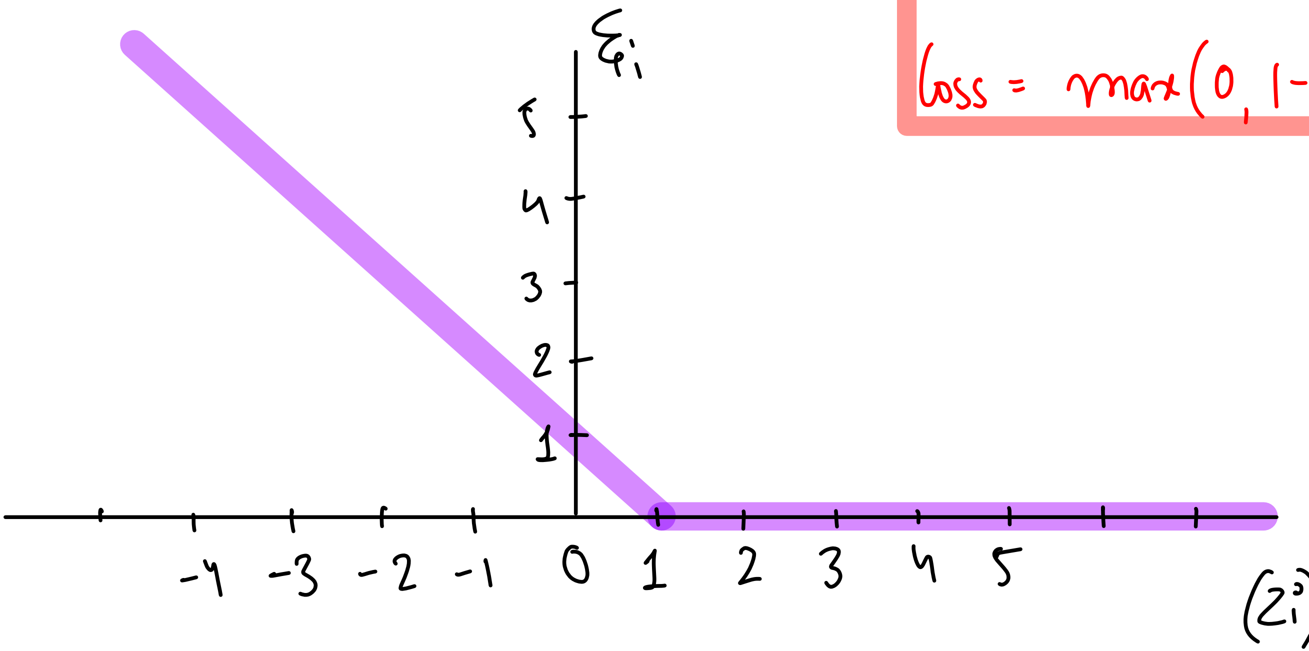
$\xi_4 = 0$



$$y_i(w^T x + b) = z_i$$

$z_i = 1 - \varepsilon_i \Rightarrow \varepsilon_i = 1 - z_i$

$\boxed{\underline{\text{HINGE LOSS}} \quad \text{loss} = \max(0, 1 - z_i)}$



$y_i(w^T x + b)$

① $\varepsilon_i = 0$

② $\underbrace{(w^T x + b) y_i}_{} \geq 1 - \varepsilon_i$

$z_i \geq 1 - \varepsilon_i$

$\varepsilon_i \geq 1 - z_i$

log Reg

$\{+1, 0\}$ — +ve, −ve

Log loss

$$y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i)$$

actual predicted.

$\{+1, -1\}$

$$\text{Loss} = \sum_{i=1}^{n} \log\left(1 + \exp\left(-y_i(w^T x + b)\right)\right)$$

$$\sum \log\left(1 + e^{-y_i(w^T x + b)}\right)$$

Svm

$\{+1, -1\}$

+ve −ve

Loss ⇒ HINGE LOSS

$$\max(0, 1 - z_i) \qquad z_i = (w^T x + b)\, y_i$$

$$\sum_{i=1}^{n} \epsilon_i \Rightarrow \max(0, 1 - z_i)$$

primal

$$\arg\min_{w} \frac{\|w\|}{2} + \frac{C}{n}\sum_{i=1}^{n} \varepsilon_i \quad \text{s.t} \quad y_i(w^T x + b) \geq 1 - \varepsilon_i \quad \forall\, 1 \to n$$

$w^*, b^* \Rightarrow$ optimal value

PRIMAL DUAL EQUIVALENCE

Dual

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

$$\text{s.t.} \quad 0 \leq \alpha_i \leq C, \quad \sum_{i=1}^{n} \alpha_i y_i = 0$$

1. $x_i \longrightarrow \alpha_i$

2. All x_i's occur in pair in the form $x_i^T x_j$

3. $$f(x_q) = \sum_{i=1}^{n} \alpha_i y_i x_i^T x_q$$

10000

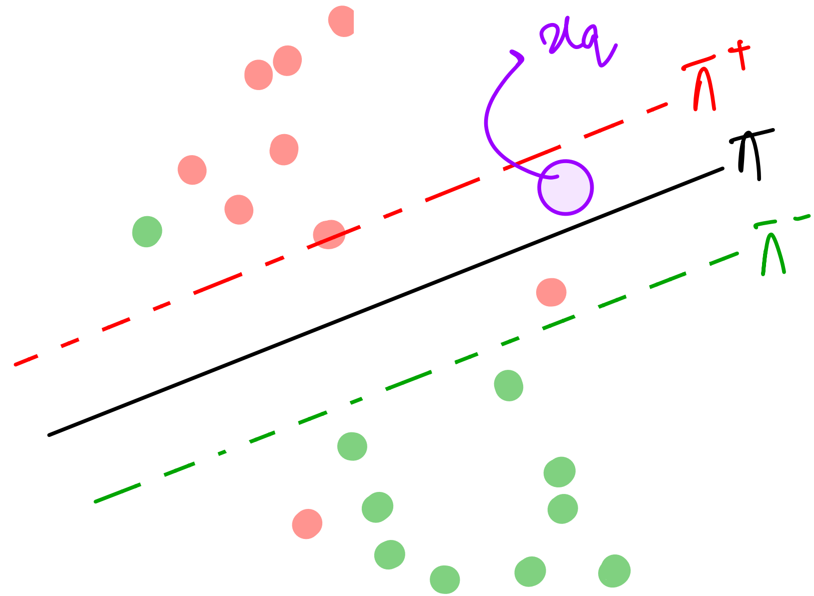
30 Support

$\alpha > 0$

4. $\alpha = 0$ For all non Support Vector

$\alpha > 0$ For all the support vector.

		w_1	w_2	w_3	
		f_1	f_2	f_3	y
α_1	x_1				
α_2	x_2				
α_3	x_3				
α_4	x_4				
α_5	x_5				



$$w_1 f_1 + w_2 f_2 + w_3 f_3 + b = 0$$

feature space ⇒ Datapoint space

$$f(x_q) = \sum_{i=1}^{n} \underbrace{\alpha_i}_{\text{learnt during training}} \underbrace{y_i}_{\text{actual label}} \underbrace{x_i}_{\text{datapoint}}^T \underbrace{x_q}_{\text{query}}$$

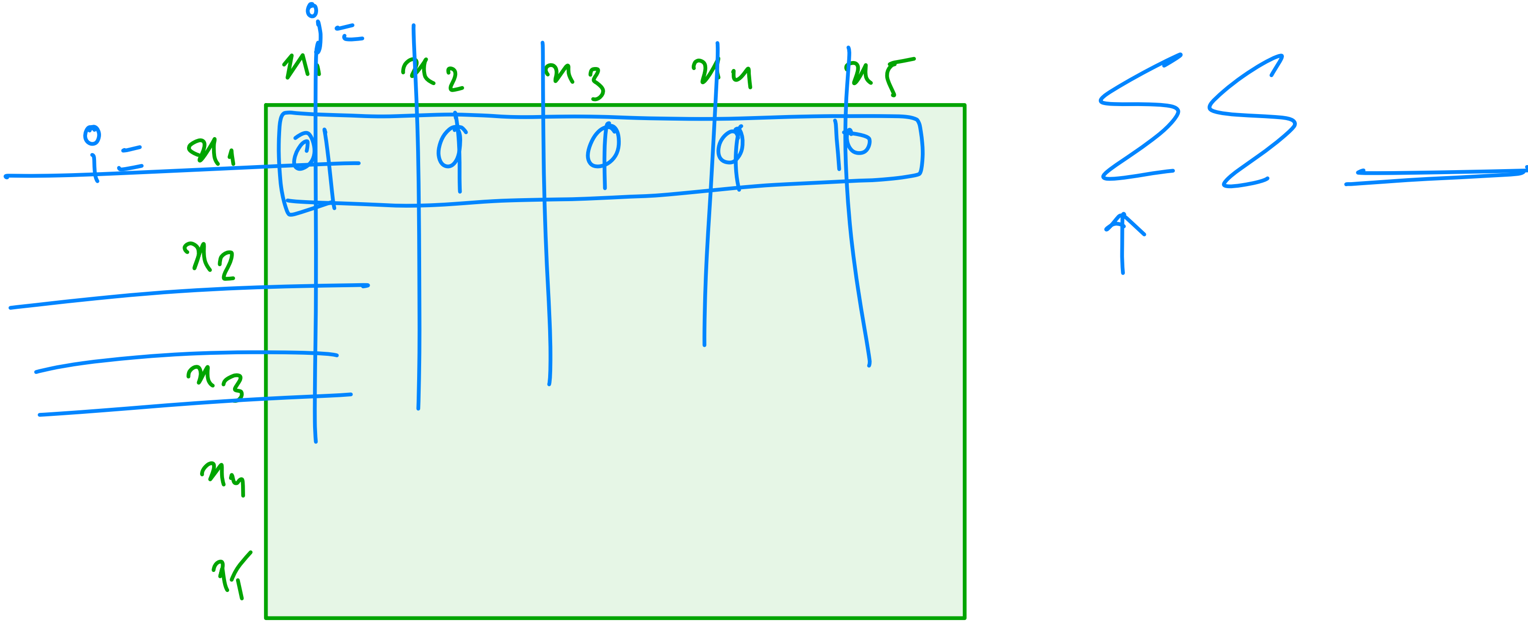
$$\underline{\underline{100}} \begin{cases} \rightarrow 97 \Rightarrow \alpha = 0 \\ \rightarrow 3 \Rightarrow \alpha > 0 \end{cases}$$

x_1
x_2
x_3

$$f(x_q) = \alpha_1 y_1 x_1^T x_q + \alpha_2 y_2 x_2^T x_q + \alpha_3 y_3 x_3^T x_q$$

$\underline{\underline{10000}} \Rightarrow 30 \Rightarrow$

9970 ✗



j =
x1
x2
x3
x4
x5
i =
x1
x2
x3
x4

dot product.

$$\underset{\alpha}{\arg\max} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \underbrace{\alpha_i \alpha_j}_{\text{parameter}} \underbrace{y_i y_j}_{\text{actual label}} \boxed{x_i^T x_j}$$

scalar value

mathematical function

representing similarity b/w x_i & x_j

Kernel Function: $K(x_i, x_j)$

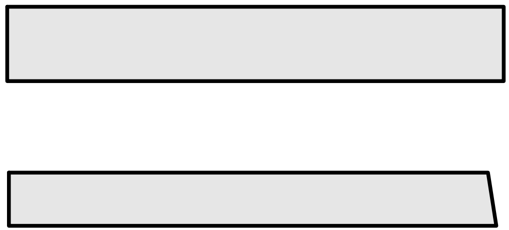
Represents similarity b/w x_i & x_j

$$x_1 = \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix}$$

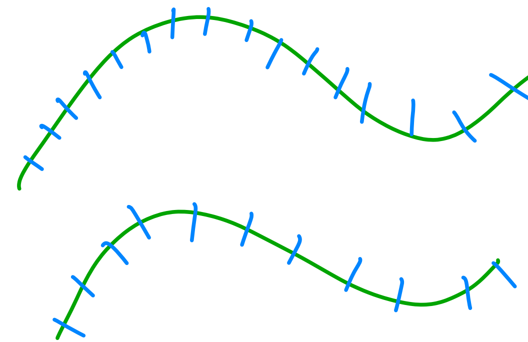
$$x_2 = \begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix}$$

Biotechnology

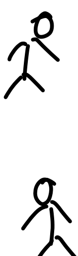
Similarity b/w DNA1 & DNA2

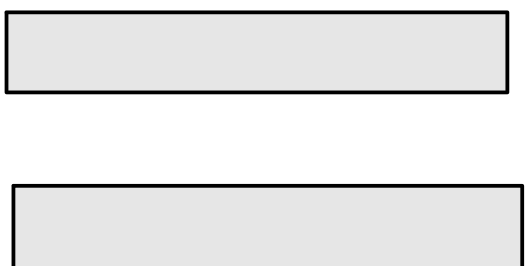
DNA_1  x_1

DNA_2 x_2



Amazon





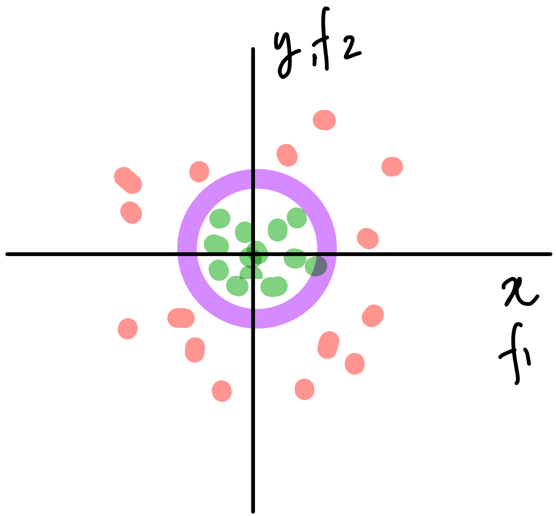
x_1

x_2

P_1 P_2

$sim(P_1, P_2)$ ✓

Kernel Function ⇒ **Polynomial Kernel.**



$x^2 + y^2 = r^2$

$x^2 + y^2 + 2gx + 2fy + c = 0$

⇒ polynomial features

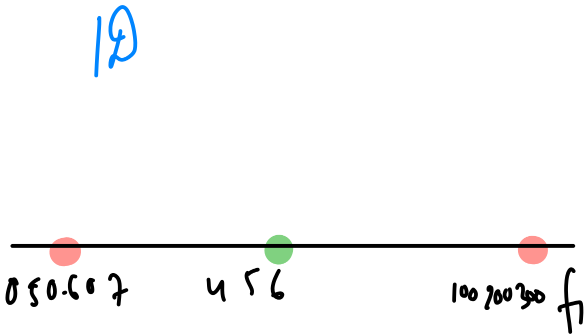
featuring engineering

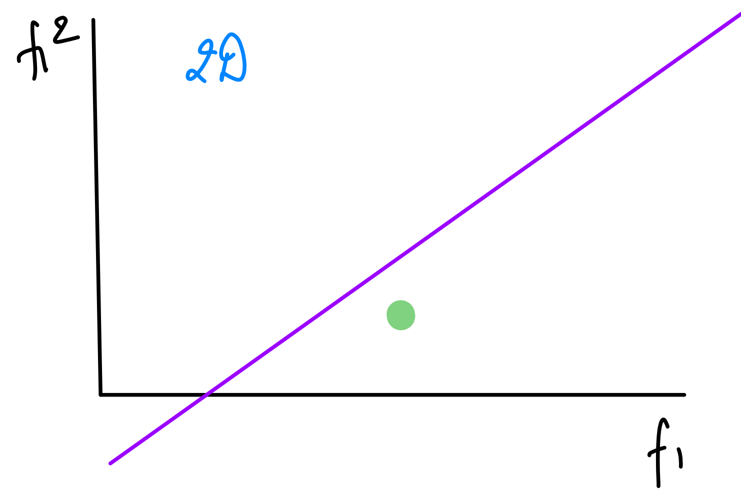
$f_1 \rightarrow f_1^2$

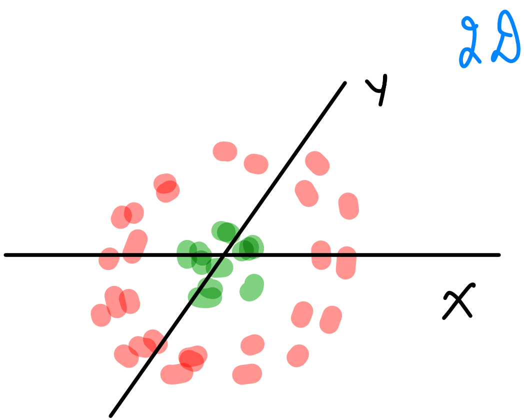
$f_2 \rightarrow f_2^2$

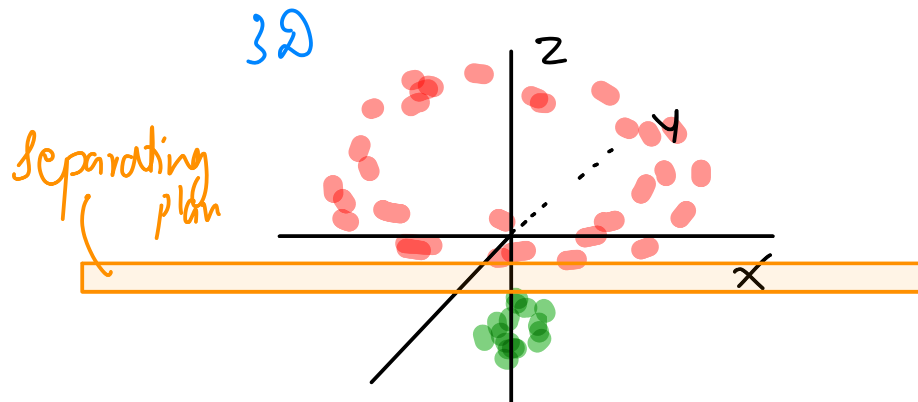
$\alpha_1 f_1^2 + \alpha_2 f_2^2$ - - - -

Eq^n of a Circle.


1D
4 5 6
100 200 300
f1


f2
2D
f1


2D
Y
X


3D
Z
Y
X
Separating plan

polynomial Kernel : $\left(c + x_1^T x_2 \right)^{n}$ = degree of polynomial

c → Const.

$c = 1$, $n = 2$

$$\boxed{K(x_1, x_2) = \left(1 + x_1^T x_2\right)^2}$$

$\left(1 + x_1^T x_2\right)^2$ Quadratic

$\left(1 + x_1^T x_2\right)^3$ Cubic

Kernelisation

d dim $\xrightarrow{\text{implicitly}}$ d' dim

$d' >> d$.

$$\dim(x_1) = 2 \qquad \dim(x_2) = 2$$

$$x_1 = \begin{bmatrix} x_{11} \\ x_{12} \end{bmatrix}$$

$$x_2 = \begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix}$$

$$K(x_1, x_2) = \left(1 + x_1^T x_2\right)^2$$

$$= \left(1 + \begin{bmatrix} x_{11} & x_{12} \end{bmatrix} \begin{bmatrix} x_{21} \\ x_{22} \end{bmatrix}\right)^2$$

$$= \left(1 + x_{11} x_{21} + x_{12} x_{22}\right)^2$$

x_1^T

$$\begin{array}{c|cc} & f_1 & f_2 \\ \hline x_1 & x_{11} & x_{12} \\ x_2 & x_{21} & x_{22} \end{array}$$

x_2^T

$$(1+a+b)^2 = a^2 + b^2 + 2ab + 2a + 2b + 1$$

$$= (1 + x_{11}x_{21} + x_{12}x_{22})^2$$

$$= \left(x_{11}^2 x_{21}^2 + x_{12}^2 x_{22}^2 + 2 x_{11}x_{21}x_{22}x_{12} + 2 x_{11} x_{21} + 2 x_{12}x_{22} + 1\right)$$

$$x_1' = [\, x_{11}^2 ,\ x_{12}^2 ,\ \sqrt{2}\, x_{11} x_{12} ,\ \sqrt{2}\, x_{11} ,\ \sqrt{2}\, x_{12} ,\ 1 \,]$$

$$x_2' = [\, x_{21}^2 ,\ x_{22}^2 ,\ \sqrt{2}\, x_{21} x_{22} ,\ \sqrt{2}\, x_{21} ,\ \sqrt{2}\, x_{22} ,\ 1 \,]$$

dot product $(x_1', x_2') \Rightarrow K(x_1, x_2)$

2 dims $\xrightarrow{\text{implicitly}}$ 6 dims

KERNEL TRICK

(x_1, x_2) (x_1', x_2')

2D 6d

<u>Same Result</u>

$\alpha = 0$ Non Support

$\alpha > 0$ Support

Quiz time!

Quiz Ended!

Which of the following are support vectors?

21 users have participated

A	Points which are within the margin	19%
B	Points which lie on +ve/-ve hyperplane ($π^+/π^-$)	38%
C	Points which are misclassified	5%
D	All of the above	38%

Quiz time!

Quiz Ended!

We have 100 datapoints out of which 5 are Support Vectors, then which is True:

20 users have participated

A	$\alpha > 0$ for 95 datapoints	20%
B	$\alpha < 0$ for 95 datapoints	5%
C	$\alpha = 0$ for 5 datapoints	10%
D	$\alpha > 0$ for 5 datapoints	65%

What of the following statement(s) is/are true about Kernel in SVM? Statement 1: Kernel function map low dimensional data to high dimensional space

Statement 2: It's a similarity function

19 users have participated

A	Statement 1	26%
B	Statement 2	0%
C	Statement 1 and 2	74%
D	None of the above	0%