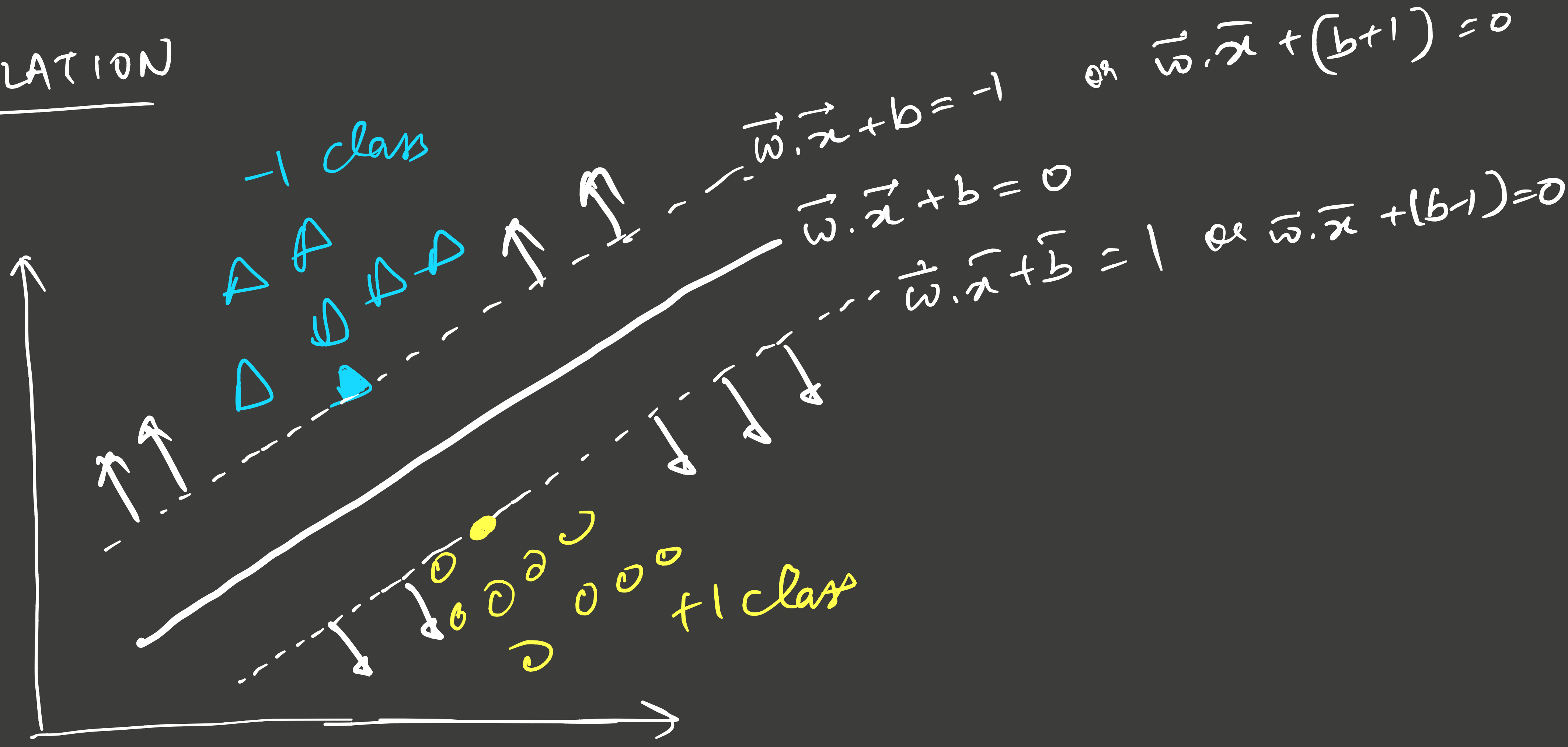
FORMULATION



$$\text{Margin} = \frac{(b+1)-(b-1)}{\|\vec{w}\|} = \frac{2}{\|\vec{w}\|}$$

mathematical convenience.

Goal: Maximize margin = Minimize $\|\vec{w}\|$ or $\frac{1}{2}\|\vec{w}\|$

s.t. correctly labelling other points.

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

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

PRIMAL FORMULATION

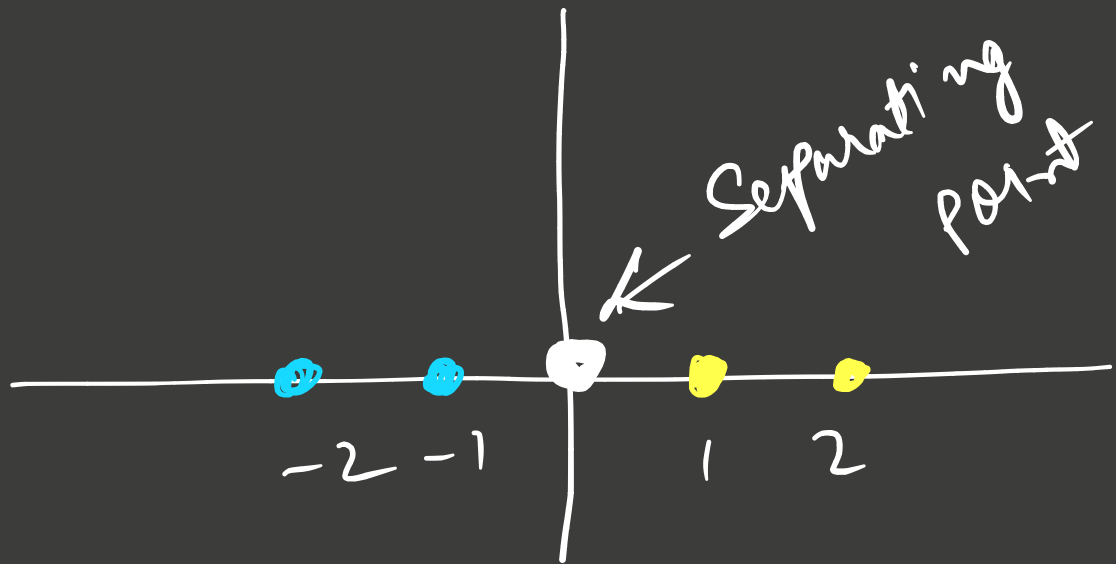
$$\boxed{\begin{aligned} &\text{Minimize} \quad \frac{1}{2}\|\vec{w}\|^2 \\ &\text{s.t.} \quad y_i(\vec{w}\cdot\vec{x_i}+b) \geq 1 \quad \forall i \end{aligned}}$$

Q) what is $\|\vec{w}\|$

$$\vec{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ \vdots \end{bmatrix} \qquad \|\vec{w}\| = \sqrt{w^T w}$$

$$= \sqrt{[w_1 \; w_2 \ldots] \begin{bmatrix} w_1 \\ w_2 \\ \vdots \end{bmatrix}}$$

SIMPLE EXAMPLE (1D)



4 points

x_1	y
1	1
2	1
-1	-1
-2	-1

Separating hyperplane: $w_1 x_1 + b = 0$

$y_i(w_1 x_1 + b) \geq 1$

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

$1(2w_1 + b) \geq 1 \quad \ldots (2)$

$-1(-w_1 + b) \geq 1 \quad \ldots (3)$

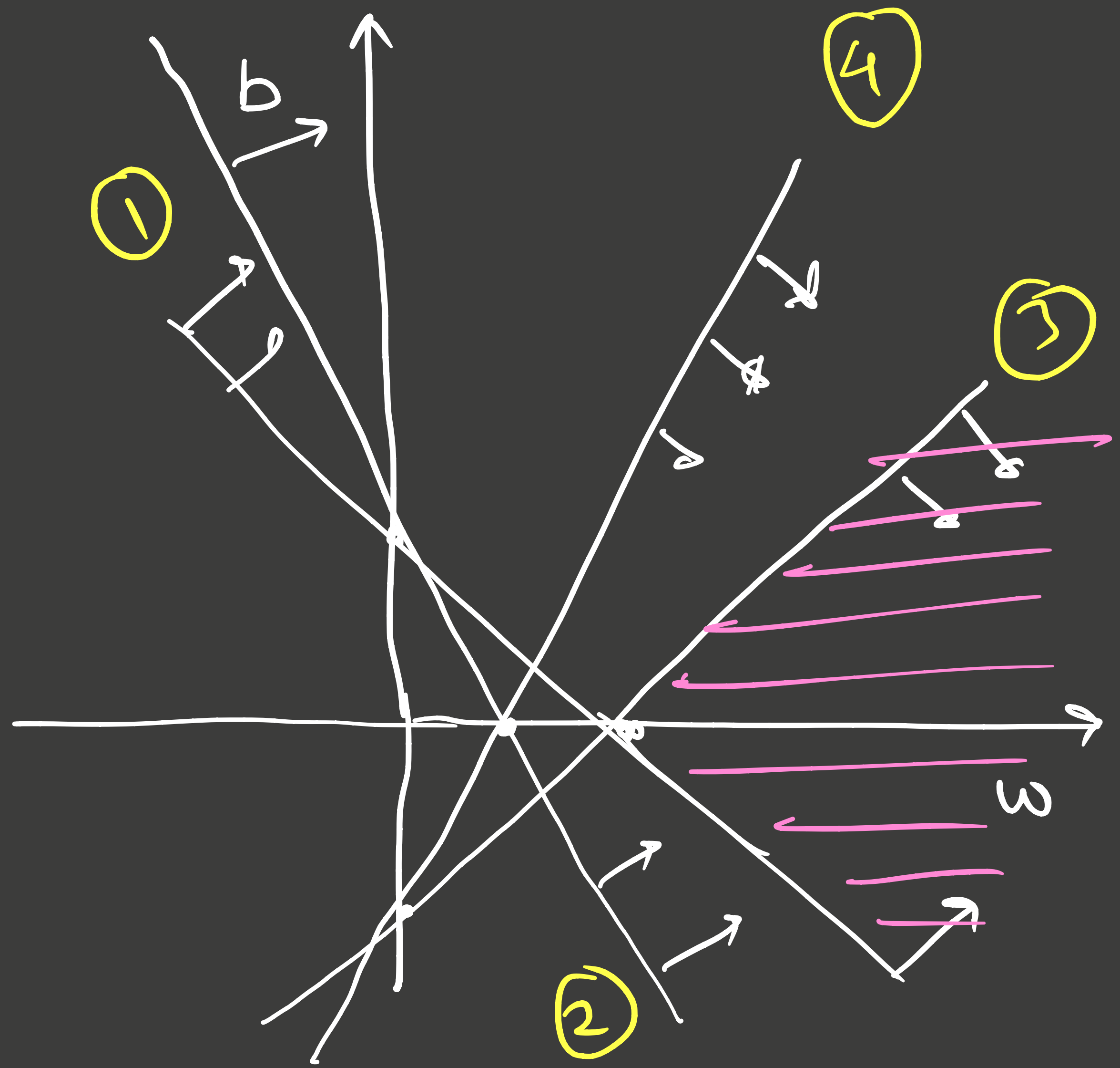
$-1(-2w_1 + b) \geq 1 \quad \ldots (4)$

$$1(w_1 + b) \geq 1 \quad \cdots (1)$$

$$1(2w_1 + b) \geq 1 \quad \cdots (2)$$

$$-1(-w_1 + b) \geq 1 \quad \cdots (3) \Rightarrow w_1 - b \geq 1$$

$$-1(-2w_1 + b) \geq 1 \quad \cdots (4) \Rightarrow 2w_1 - b \geq 1$$



$$w_{min} = 1$$

$$b = 0$$

$$\vec{w} \cdot \bar{x} + b = 0$$

or $x = 0$

Minimum w_1 satisfying constraints is

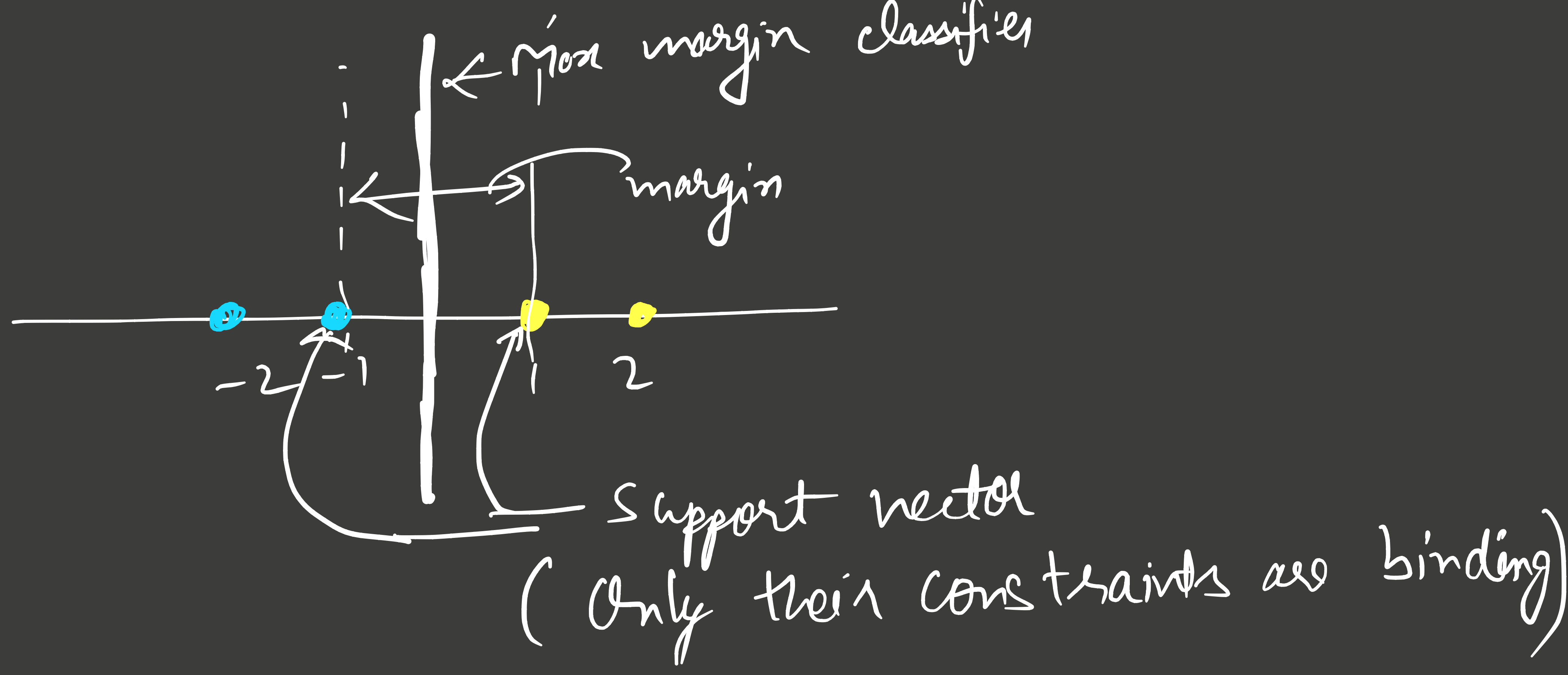
$$w_1 = 1$$

&

correspondingly $b = 0$

$\therefore$ Max. margin classifier is $1 \cdot x + 0 = 0$

or $x = 0$



PRIMAL FORMULATION IS A QUADRATIC PROGRAM (QP)

Generally;

Minimize Quadratic (x)

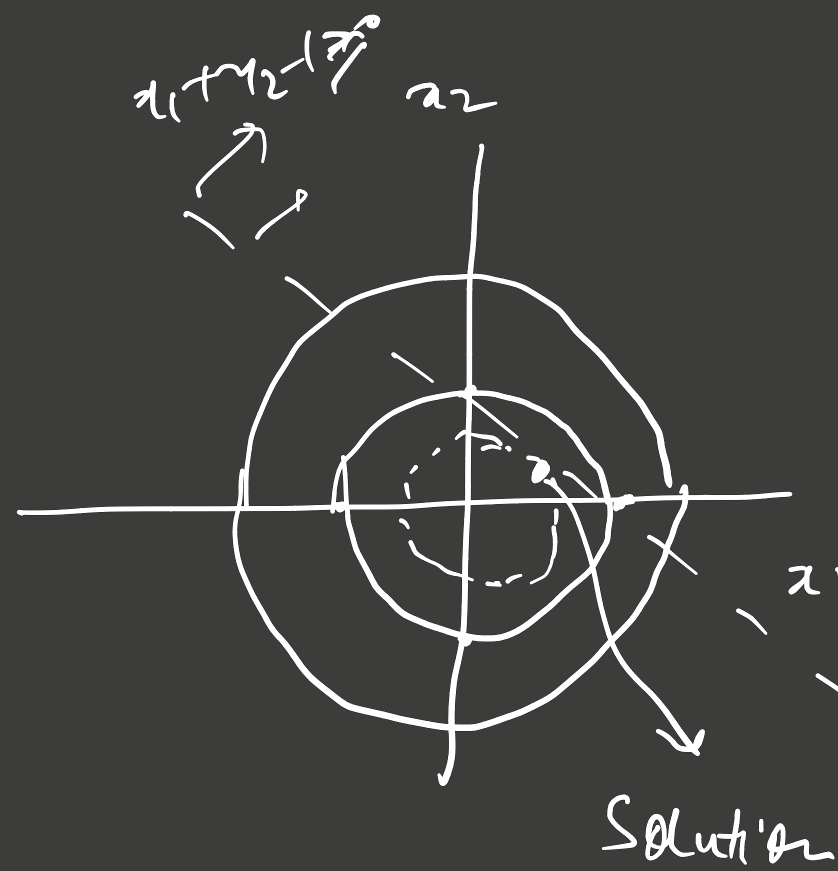
s.t. linear (x)

eg. $\bar{x} = (x_1, x_2)$

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

$$\text{s.t.} \quad x_1 + x_2 - 1 \geq 0$$

Sol^n: $x_1 = x_2 = 1/2$



PRIMAL $\longleftrightarrow$ DUAL CONVERSION

(USE LAGRANGIAN MULTIPLIERS)

$$\text{Min } \frac{1}{2} \|\bar{w}\|^2 \quad \text{s.t.} \quad y_i(\vec{w} \cdot \vec{x}_i + b) - 1 \geq 0$$

$$\text{or } \frac{1}{2} \sum_{i=1}^{d} w_i^2$$

Let $\vec{w} \in R^d$; $i \in \{1, .. N\}$ TRAINING EXAMPLES

$$L(\vec{w}, b, \alpha_1, \alpha_2 .. \alpha_N) = \frac{1}{2} \sum_{i=1}^{d} w_i^2 - \sum_{i=1}^{N} \alpha_i (y_i (\vec{w} \cdot \vec{x}_i + b) - 1)$$

$$\alpha_1, \alpha_2 \dots \alpha_N \geq 0$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0 \quad \dots (1)$$

$$\frac{\partial L}{\partial \vec{w}} = 0 \Rightarrow \quad \vec{w} - \sum_{i=1}^{N} \alpha_i y_i \vec{x_i} = 0$$

$$\Rightarrow \quad \vec{w} = \sum_{i=1}^{N} \alpha_i y_i \vec{x_i} \quad \cdots \quad (2)$$

Using (1) & (2);

$$L(\alpha_1 \cdots \alpha_N) = \frac{1}{2} \|\vec{w}\|^2 - \sum_{i=1}^{N} \alpha_i \left(y_i (\vec{w} \cdot \vec{x_i} + b) - 1 \right)$$

$$= \frac{1}{2} \|\vec{w}\|^2 - \sum_{i=1}^{N} \alpha_i y_i \vec{w} \cdot \vec{x_i} - \overset{0}{\cancel{\sum_{i=1}^{N} \alpha_i y_i b}} + \sum_{i=1}^{N} \alpha_i$$

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

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

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

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

s.t. $\quad y_i (\vec{w} \cdot x_i + b) \geq 1$

s.t. $\quad \sum_{i=1}^{N} \alpha_i y_i = 0$

$\alpha_i \geq 0 \quad \forall i \in \{1, .., N\}$

FINALLY let's add KKT complementary slackness

$$\alpha_i (y_i (\vec{w} \cdot \vec{x_i} + b) - 1) = 0 \quad \forall i$$

Q. we have

$$\alpha_i \left(y_i (\vec{w} \cdot \vec{x}_i + b) - 1 \right) = 0 \quad \forall i$$

What is α_i for support vector points?

Ans). For support vectors:

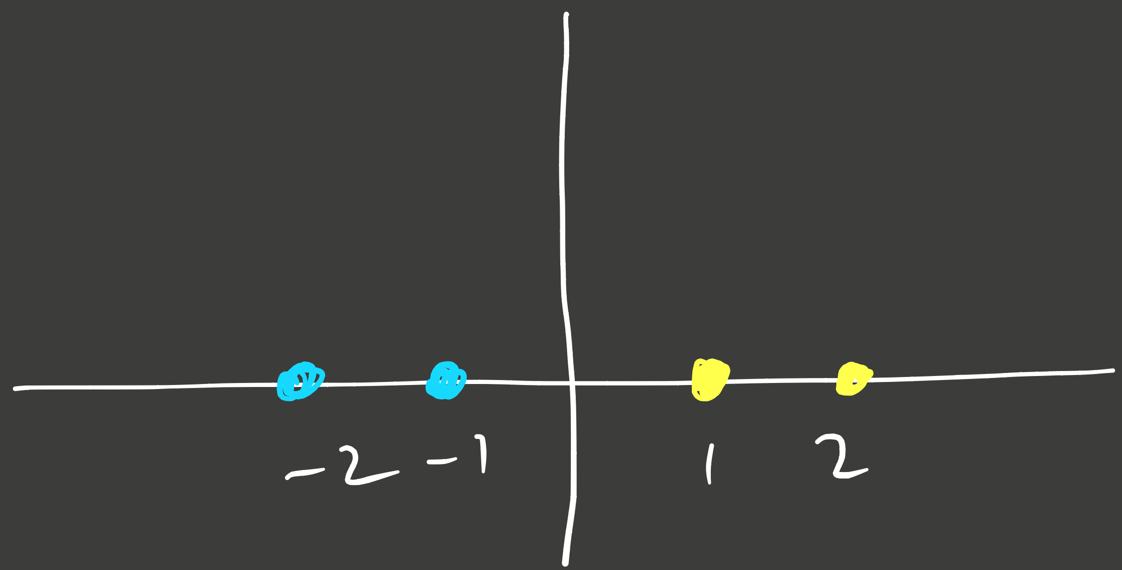
$$\bar{w} \cdot \bar{x}_i + b = 1 \quad (+\text{ve class})$$
$$= -1 \quad (-\text{ve class})$$

or $y_i (\bar{w} \cdot \bar{x}_i + b) - 1) = 0$ for $i = \{\text{S.V. points}\}$

$\therefore$ α_i where $i \in$ S.V. points is $\neq 0$

FOR ALL NON-SUPPORT VECTOR POINTS; $\alpha_i = 0$

Revisiting the simple example (1-D) in dual.



4 points

x_1	y
1	1
2	1
-1	-1
-2	-1

$$L(\alpha) = \sum_{i=1}^{4} \alpha_i - \frac{1}{2} \sum_{i=1}^{4} \sum_{j=1}^{4} \alpha_i \alpha_j y_i y_j \vec{x_i} \vec{x_j}$$

s.t. $\alpha_i \geq 0$

$$\sum \alpha_i y_i = 0$$

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

$$L(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4$$

$$-\frac{1}{2}\left\{ \alpha_1 \alpha_1 * (1*1) * (1*1) \right.$$

$$+$$

$$\alpha_1 \alpha_2 * (1*1) * (1*2)$$

$$+$$

$$\alpha_1 \alpha_3 * (1*-1) * (1*-1)$$

$$+$$

$$\cdots$$

$$\cdots$$

$$+$$

$$\cdots$$

$$\left. \alpha_4 \alpha_4 * (-1*-1) * (-2*-2) \right\}$$

How to solve? QP solver

FOR TRIVIAL EXAMPLE,

BY SYMMETRY, $\alpha_2 = \alpha_4 = \alpha$ (say)

& (ALSO $\sum y_i \alpha_i = 0$)

$\alpha_1 = \alpha_3 = 0$

MAXIMIZE $(\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4)$

$$-\frac{1}{2}\Big[\alpha_2\alpha_4(1)(-1)(2)(-2)$$

$$+ \alpha_4\alpha_2(-1)(1)(-2)(2)\Big]$$

$$\underset{\alpha}{\text{MAXIMIZE}} \quad 2\alpha - \frac{1}{2}\left[4\alpha^2 + 4\alpha^2\right]$$

$$\frac{\partial}{\partial \alpha}\left(2\alpha - 4\alpha^2\right) = 0 \Rightarrow 2 - 8\alpha = 0$$

$$\Rightarrow \alpha = 1/4$$

$$\frac{\partial}{\partial \alpha^2}\left(2\alpha - 4\alpha^2\right) = -8$$

$\therefore$ Maxima.

$$\therefore \alpha_1 = 0;\ \alpha_2 = 1/4;\ \alpha_3 = 0;\ \alpha_4 = 1/4$$

$$\vec{w} = \sum_{i=1}^{N} \alpha_i y_i \vec{x}_i = 0 \times 1 \times 1 + \frac{1}{4} \times 1 \times 2 + 0 \times -1 \times -1 + 1/4 \times -1 \times -2$$

$$= 2/4 + 2/4 = 1$$

FINDING 'b'

FOR SUPPORT VECTORS WE HAVE

$$\alpha_i = 0$$

&

$$y_i(\vec{w}\cdot\vec{x_i} + b) - 1 = 0 \qquad i \in S.V.$$

or $y_i(\bar{w}\cdot\bar{x_i} + b) = 1$

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

or $\bar{w}, \bar{x_i} + b = y_i \qquad (\because y_i^2 = 1)$

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

In practice $b = \dfrac{1}{N_{SV}} \displaystyle\sum_{i=1}^{N_{SV}} (y_i - \bar{w}\cdot\vec{x_i})$

So,

$$b = \frac{1}{2}\left\{\left(1 - (1)(1)\right) + \left(-1 - (1)(-1)\right)\right\}$$

$$= \frac{1}{2}\left\{0 + 0\right\} = 0$$

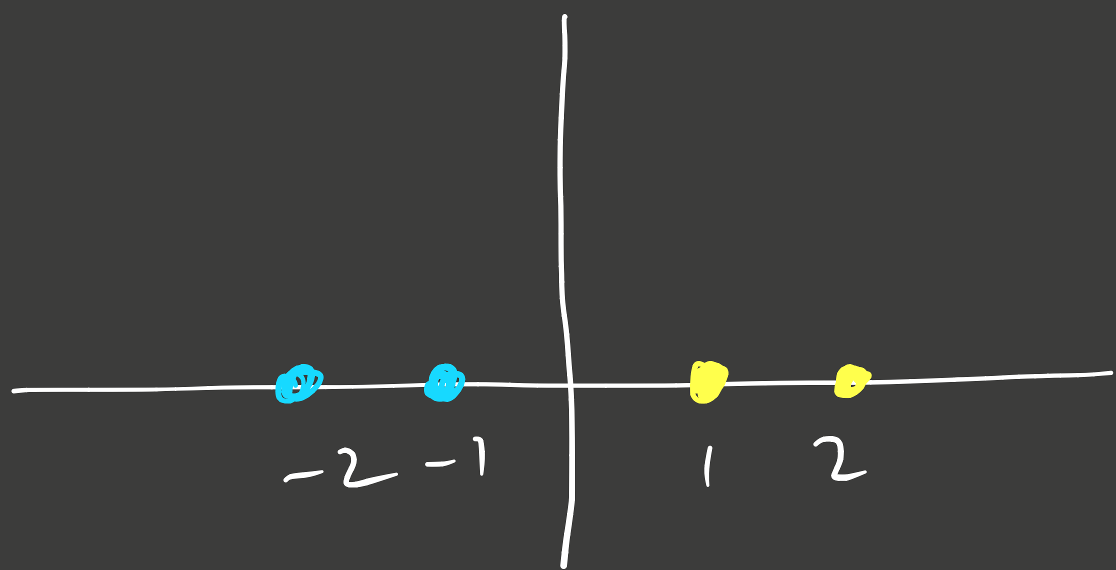
$\therefore \quad w = 1$

&

$b = 0$

MAKING PREDICTIONS

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



For $x_{\text{TEST}} = 3$; $\hat{y}(3) = \text{SIGN}(1 * 3 + 0)$

$= +$

$=$ +ve class

ALTERNATIVELY,

$$\hat{y}(x_{TEST}) = \text{SIGN}(\vec{w} \cdot \vec{x}_{TEST} + b)$$

$$= \text{SIGN}\left(\sum_{j=1}^{N_{SV}} \alpha_j \, y_j \, \vec{x}_j \cdot \vec{x}_{Test} + b\right)$$

In our example

$\alpha_1 = 0; \; \alpha_2 = 1/4; \; \alpha_3 = 0; \; \alpha_4 = 1/4$

$$\hat{y}(3) = \text{SIGN}\left(\frac{1}{4} * 1 * (2*3) + 0 + \frac{1}{4} * (-1) * (-2*3) + 0\right)$$

$$= \text{SIGN}\left(\frac{12}{4}\right) = \text{SIGN}(3) = \text{+ve class}$$

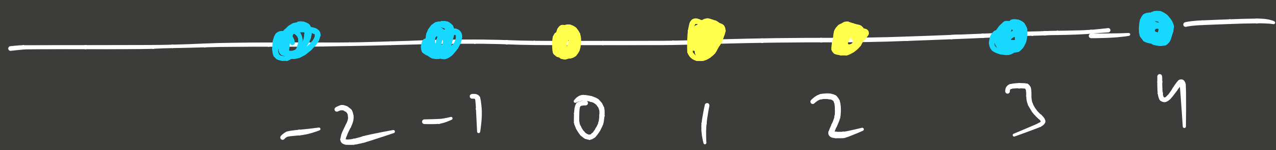
PRIMAL WAS QP, DUAL IS QP,

WHY BOTHER CONVERTING TO DUAL?

JUST WAIT FOR FEW MORE MINS

ANSWER: "KERNEL TRICK")

NON-LINEARLY SEPARABLE DATA

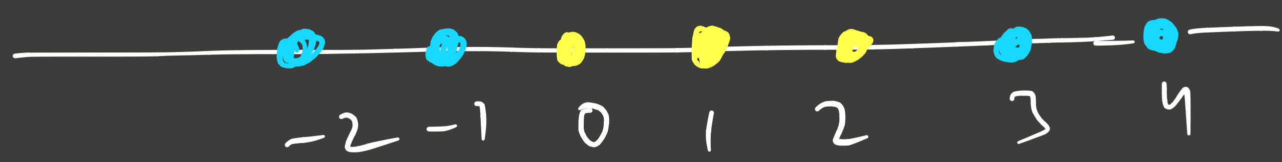


Data not separate in R

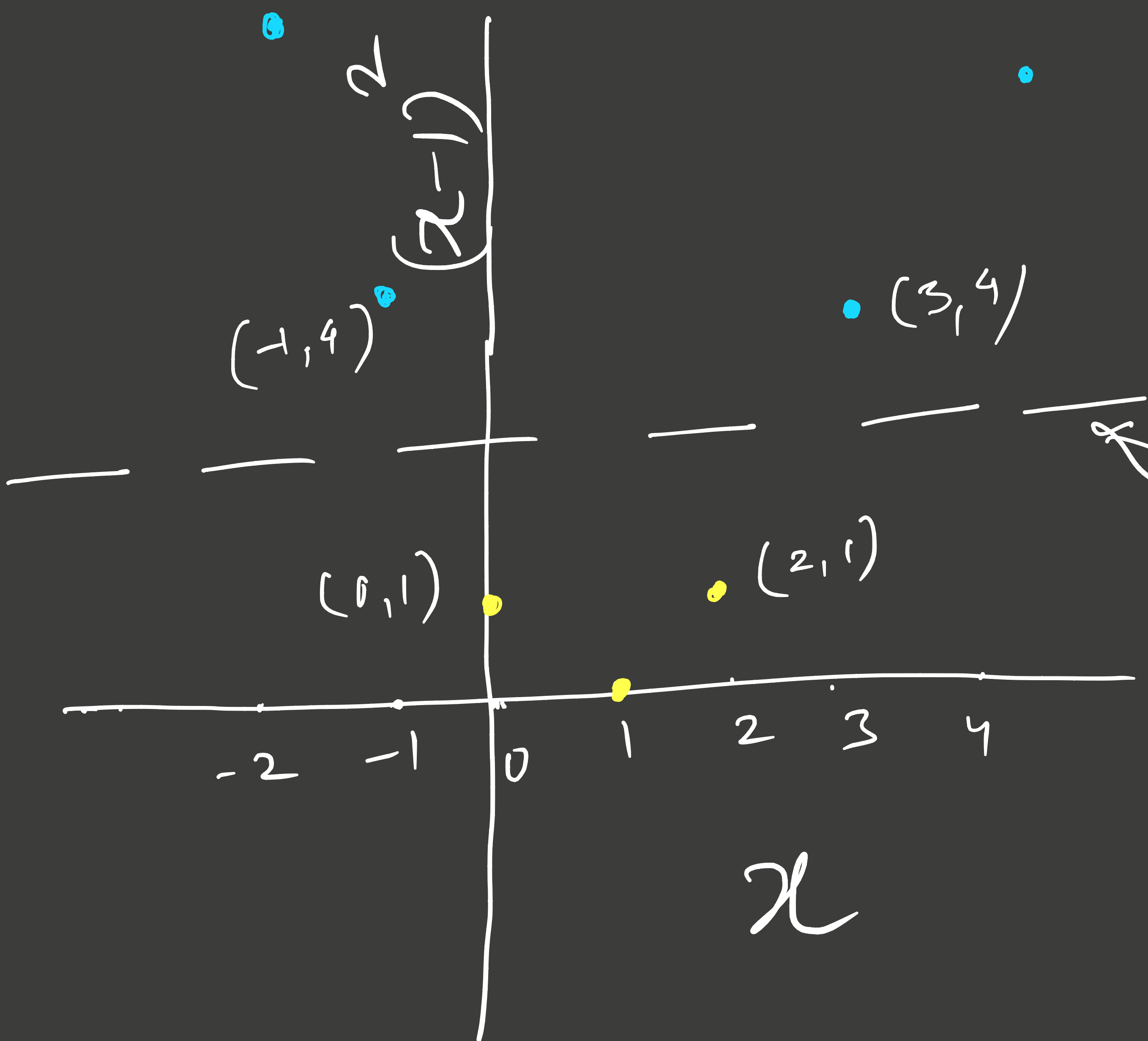
Can we still use SVM?

Yes!

How: Project data to a higher dimensional space.

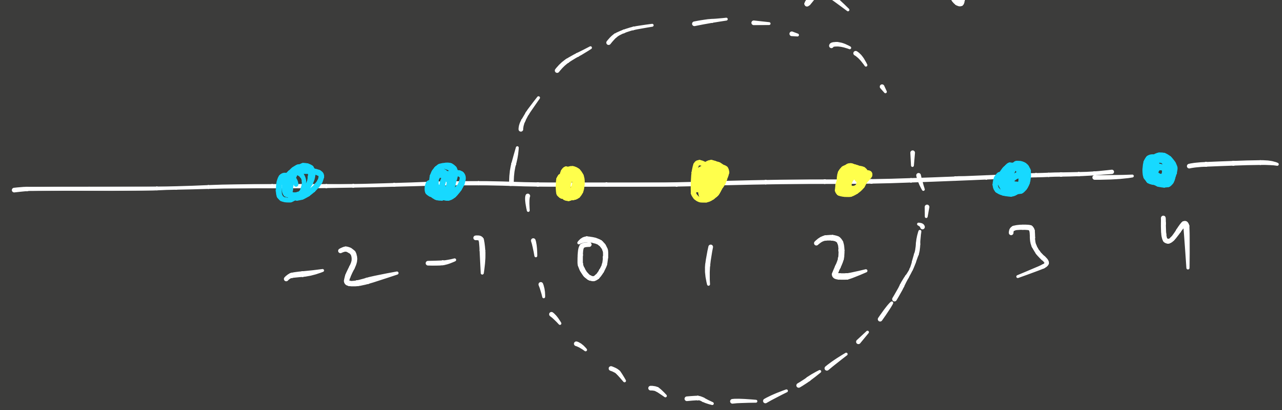


original data in R



Transformed data in R^2

Max margin separating hyperplane



separating plane in original dimension

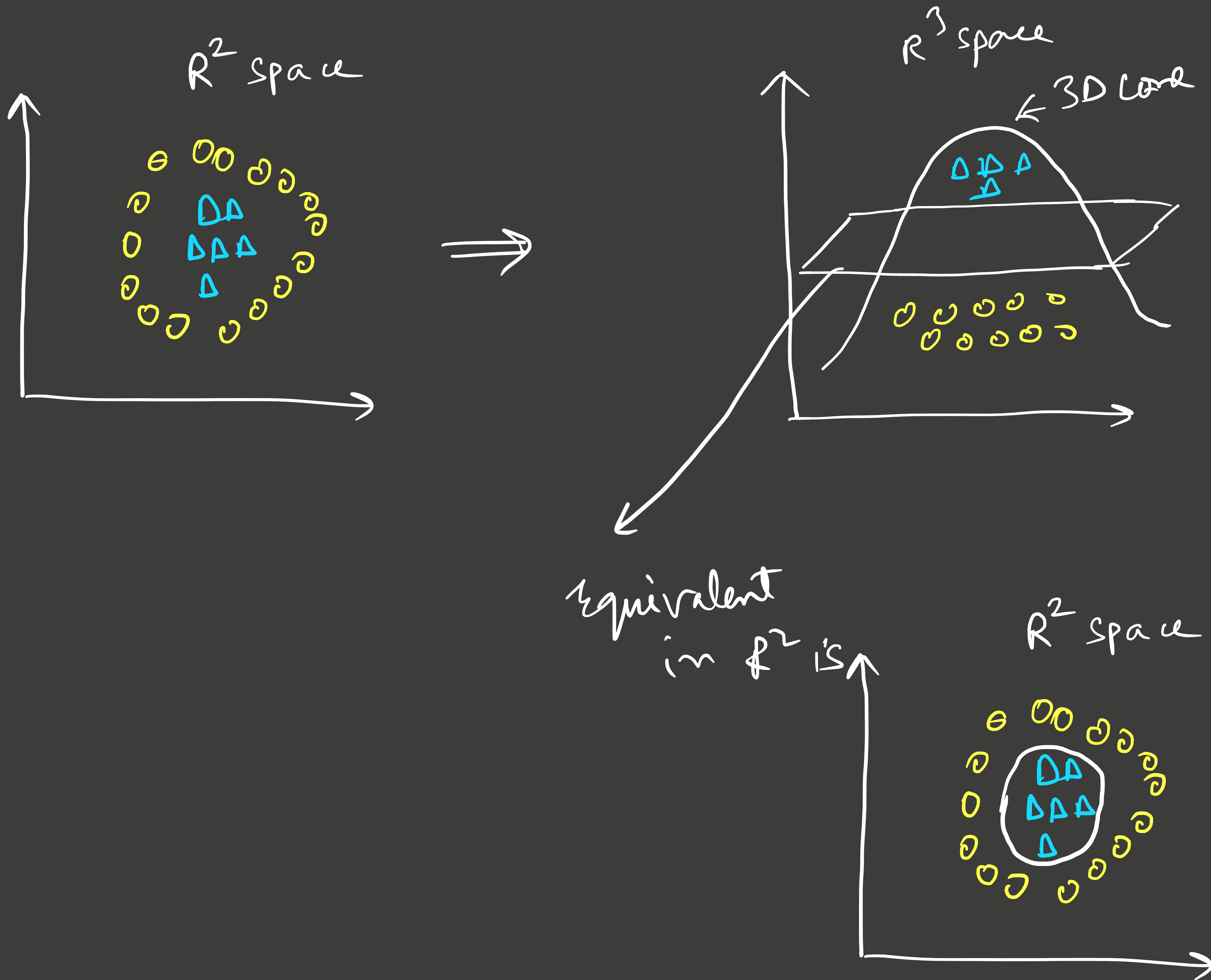
$$(x-1)^2 = \left(\sqrt{\frac{5}{2}}\right)^2$$

Circle :

Center $(x = 1)$

Radius $\sqrt{5/2}$

ANOTHER EXAMPLE TRANSFORMATION



Projection / Transformation Function

$$\phi : R^d \longrightarrow R^D$$

where d = original dimension

D = New dimension

In our example;

$d = 1$; $D = 2$

Linear SVM

MAXIMIZE

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

s.t.

CONSTRAINTS

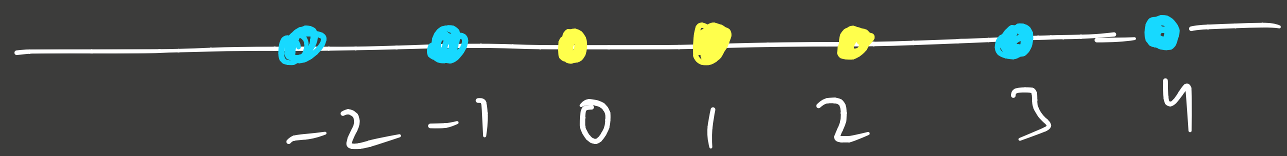
$\Downarrow$

TRANSFORMATION (ϕ)

$\Downarrow$

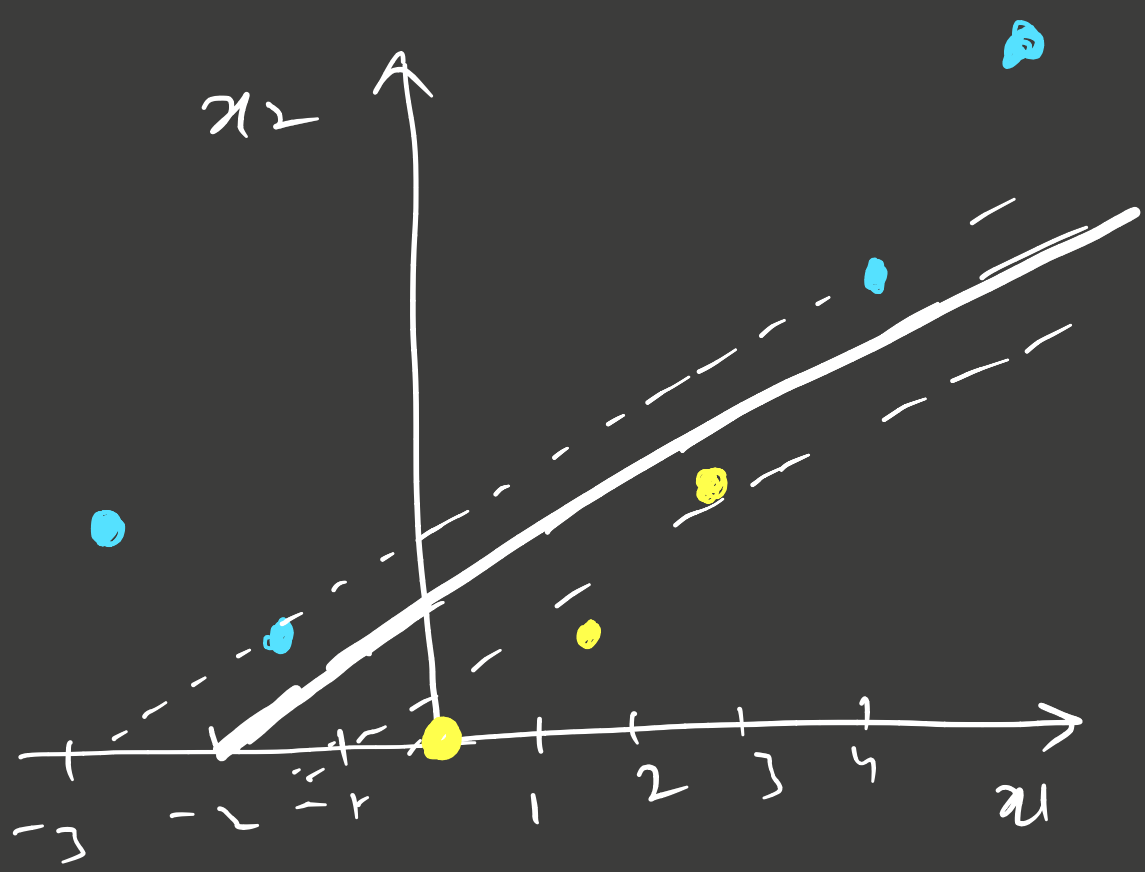
$$L(\alpha) = \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j \phi(\vec{x_i}) \cdot \phi(\vec{x_j})$$

TRIVIAL EXAMPLE (again)



Original Data $(x) \in R$

Transformed Data $\left(\phi(x) = \langle \sqrt{2}x, x^2 \rangle \right)$



Steps

① Compute $\phi(x)$ for each point

$$\phi : R^d \rightarrow R^D$$

② Compute dot products over R^d space

③ If $D >> d$

Both steps are expensive!

KERNEL TRICK

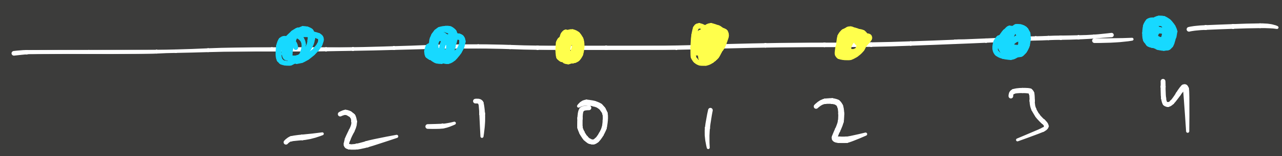
Can we compute $K(\bar{x}_i, \bar{x}_j)$

s.t.

$$K(\bar{x}_i, \bar{x}_j) = \phi(\bar{x}_i) \cdot \phi(\bar{x}_j)$$

some func^n of dot product in original dimension

Dot product in high dimensions (after transformation)



$$\phi(x) = \langle \sqrt{2}\, x, \; x^2 \rangle$$

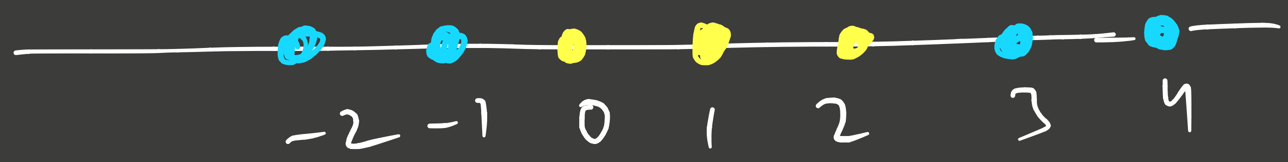
$$K(x_i, x_j) = (1 + x_i x_j)^2 - 1$$

Dot product in lower dimension

$$= \cancel{1} + 2 x_i x_j + x_i^2 x_j^2 - \cancel{1}$$

$$= \langle \sqrt{2}\, x_i, \; x_i^2 \rangle \cdot \langle \sqrt{2}\, x_j, \; x_j^2 \rangle$$

$$= \phi(x_i) \cdot \phi(x_j)$$



Original Dataset

#	x	y
1	-2	-1
2	-1	-1
3	0	1
⋮	⋮	⋮

Transformed dataset

#	$\sqrt{2}x$	x^2	y
1	$-2\sqrt{2}$	4	-1
2	$-\sqrt{2}$	1	-1
3	0	0	1
⋮	⋮	⋮	⋮

$\phi(x_1) = \langle -2\sqrt{2}, 4) ; \phi(x_2) = \langle -\sqrt{2}, 1 \rangle$ TRANSFORMAT^N

$\phi(x_1) . \phi(x_2) = -2\sqrt{2} * -\sqrt{2} + 4 \times 1 = 8$ DOT PRODUCT IN 2D

$K(x_1, x_2) = \{1 + (-2) * (-1)\}^2 - 1$ DOT PRODUCT IN 1D

WHY DID WE USE DUAL FORM?

KERNELS AGAIN!!

PRIMAL FORM DOESN'T ALLOW FOR "KERNEL TRICK" $K(\vec{x}_1, \vec{x}_2)$ in DUAL

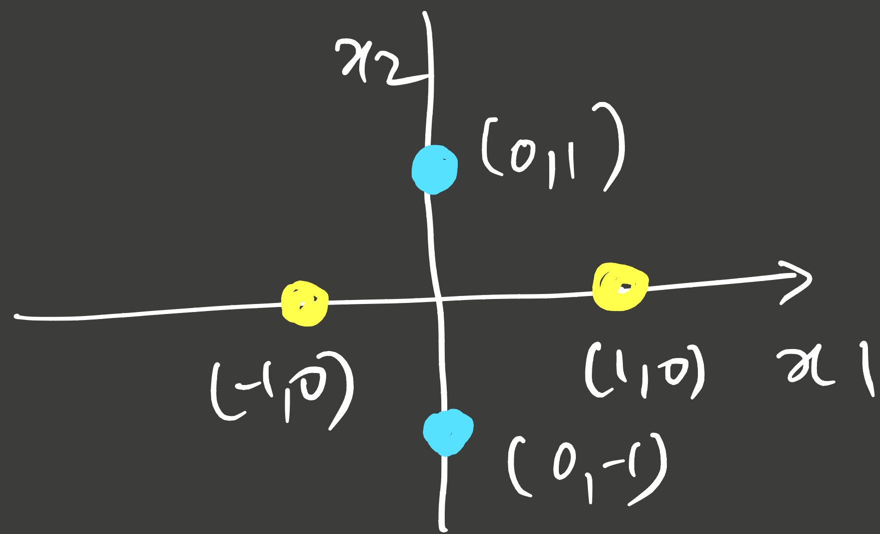
& COMPUTE $\phi(x)$ and then dot product in 'D' dimensions?

GRAM MATRIX (Positive Semi-Definite)

$$K(x_i, x_j) = (1 + x_i \cdot x_j)^2 - 1$$

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1	24	8	0	0	8	24	48
x_2	8	1	0	-1	0		
x_3	0	..	..	..	..		
x_4	0						
x_5	8						
x_6	24						
x_7	48						

ANOTHER EXAMPLE



$$K(\bar{x}, \bar{x}') = (\vec{x}^T \cdot \vec{x}')^2$$

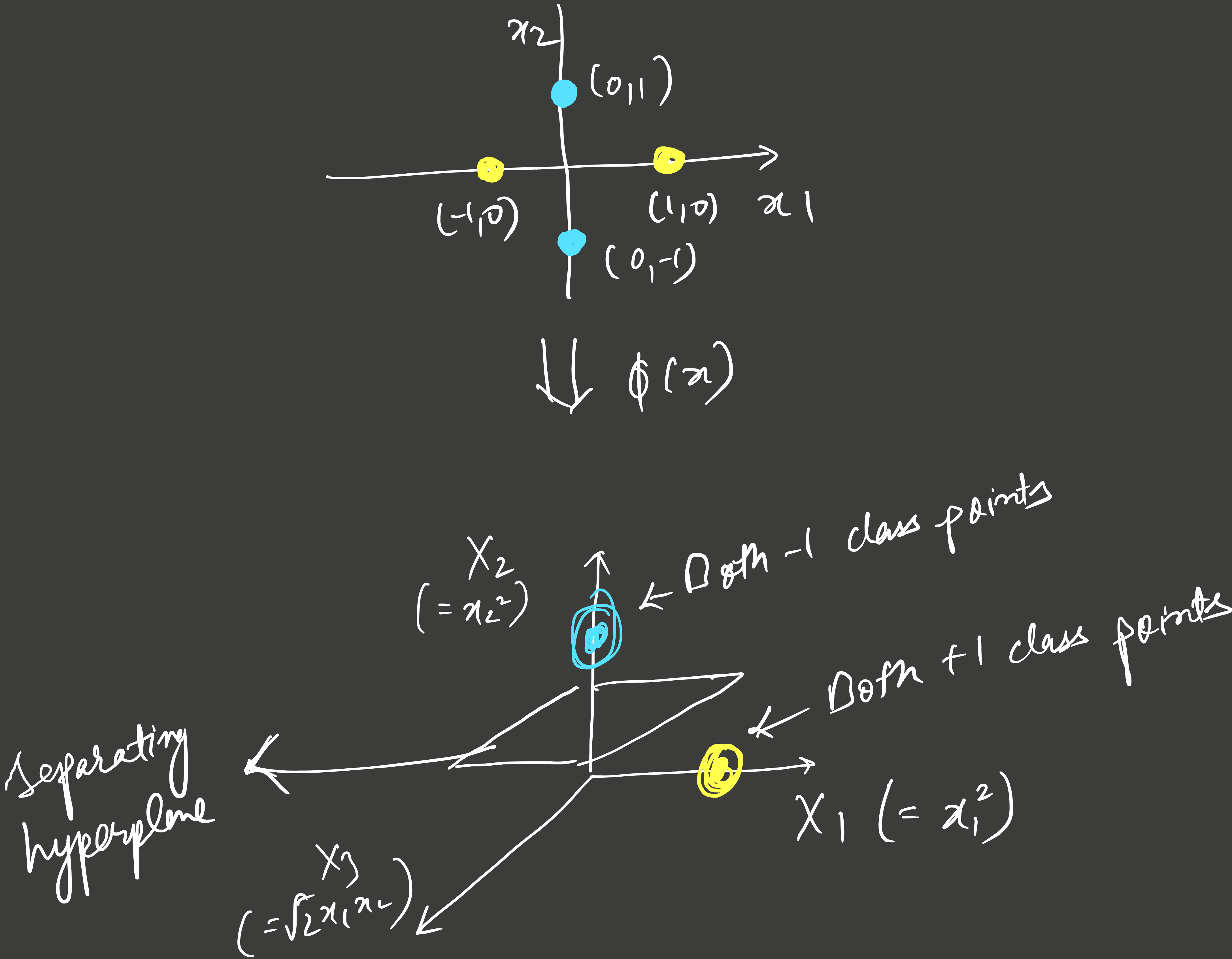
$$\bar{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \; x' = \begin{bmatrix} x_1' \\ x_2' \end{bmatrix}$$

Q: What is $\phi(x)$?

$$K(\bar{x}, \vec{x}') = \phi(\vec{x}) \cdot \phi(\vec{x}')$$

$$K(\vec{x}, \vec{x}') = \left\{ \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} x_1' \\ x_2' \end{bmatrix} \right\}^2 = (x_1 x_1' + x_2 x_2')^2$$

$$\Rightarrow \boxed{\phi(x) = \langle x_1^2, \sqrt{2} x_1 x_2, x_2^2 \rangle} = x_1^2 x_1'^2 + x_2^2 x_2'^2 + 2 x_1 x_1' x_2 x_2'$$


x_2
$(0,1)$
$(-1,0)$
$(1,0)$
x_1
$(0,-1)$
$\Downarrow \phi(x)$
X_2
$(= x_2^2)$
Both -1 class points
Both +1 class points
Separating hyperplane
X_1 $(= x_1^2)$
X_3
$(= \sqrt{2} x_1 x_2)$

SOME KERNELS

① Linear: $K(\bar{x}_1, \bar{x}_2) = \bar{x}_1 \cdot \bar{x}_2$

② Polynomial: $K(\bar{x}_1, \bar{x}_2) = (p + \bar{x}_1 \cdot \vec{x}_2)^q$

③ Gaussian: $K(\bar{x}_1, \bar{x}_2) = e^{-\gamma \|\bar{x}_1 - \bar{x}_2\|^2}$

ALSO CALLED RADIAL BASIS FUNCTION (RBF)

$$\gamma = \frac{1}{2\sigma^2}$$

Q) For $\bar{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ what space does

Kernel $K(\bar{x}, \vec{x}') = (1 + \bar{x} \cdot \vec{x}')^3$ belong to?

or

$\vec{x} \in R^2$

$\phi(\vec{x}) \in R^?$

$$K(x, z) = (1 + x_1 z_1 + x_2 z_2)^3$$

$$= \ldots$$

$$= \langle 1, x_1, x_2, x_1^2, x_2^2, x_1^2 x_2, x_1 x_2^2, x_1^3, x_2^3 \rangle$$

10 dimensional?

Q) For $\bar{x} = x$; what space does RBF kernel lie in?

$$K(x,z) = e^{-\gamma \|x-z\|^2}$$

$$= e^{-\gamma (x-z)^2}$$

Now; $e^{\alpha} = \sum_{n=0}^{\infty} \frac{\alpha^n}{n!}$

$\therefore$ $e^{-\gamma (x-z)^2}$ is ∞ dimensional!!

Q) Is SVM parametric or non-parametric?

Q) Is SVM parametric or non-parametric?

Yes and No

↓ Linear kernel / Polynomial kernel (form fixed)

↓ RBF (form changes with data)

RBF is Non Parametric

$$\hat{y}(x_{TEST}) = \text{SIGN}(\vec{w} \cdot \vec{x}_{TEST} + b)$$

$$= \text{SIGN}\left(\sum_{j=1}^{N_{SV}} \alpha_j y_j \vec{x}_j \cdot \vec{x}_{Test} + b\right)$$

$\Downarrow$ Kernelized.

$$\hat{y}(x_{Test}) = \text{SIGN}\left(\sum_{j=1}^{N} \alpha_j y_j K(\vec{x}_j, \vec{x}_{Test}) + b\right)$$

$\alpha_j = 0$ where $j \neq$ S.V.

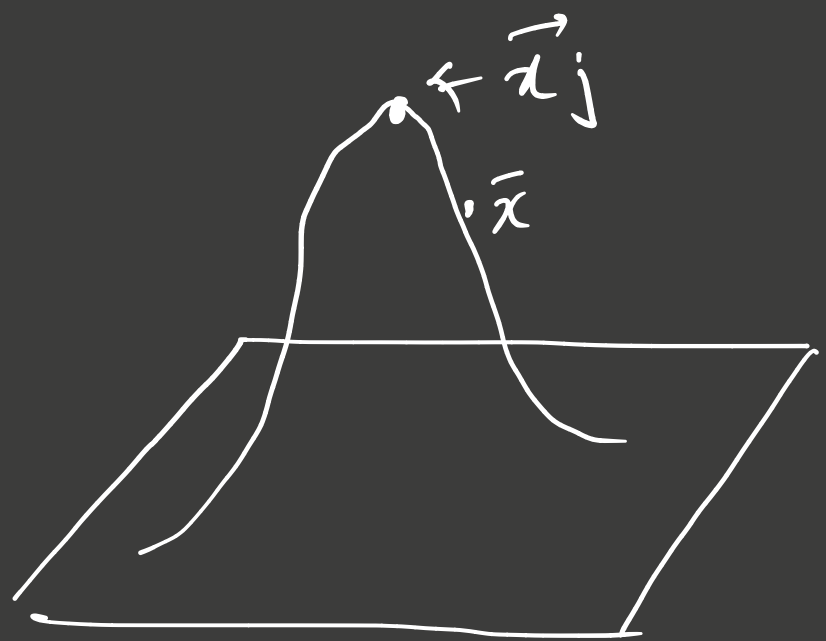
Now $K(\vec{x}_j, \vec{x}_{Test})$ for RBF is:

$$e^{-\gamma \|\underline{\underline{\vec{x}_j}} - \vec{x}_{Test}\|^2}$$

$\therefore$ Hypothesis is a $\underline{\underline{\text{function}}}$ of "All" train points.

↓

What kind of?



Closer $\vec{x}$ is to $\vec{x}_N$, more is it influencing $\underline{\underline{\hat{y}}}(\vec{x})$

← Hypothesis function

Now if we add a point to dataset

$\Downarrow$

Functional form can adapt (similar to KNN)

$\therefore$ SVM with RBF Kernel is Non-Parametric

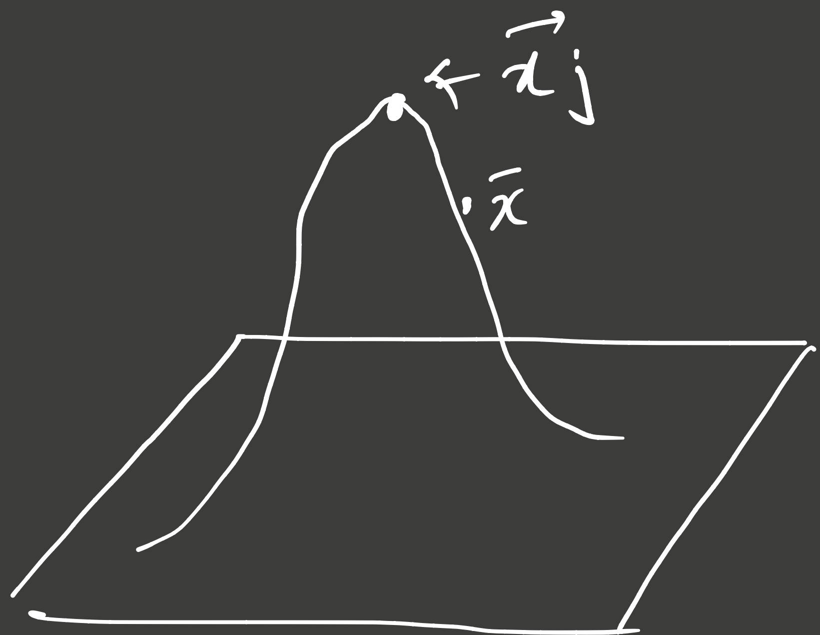
Interpretation of RBF

Radial

$$\hat{y}(x) = \text{SIGN}\left(\sum \underbrace{\alpha_i y_i}_{\text{Activation}} \underbrace{e^{-\|x - x_i\|^2}}_{\text{Basis}} + b \right)$$

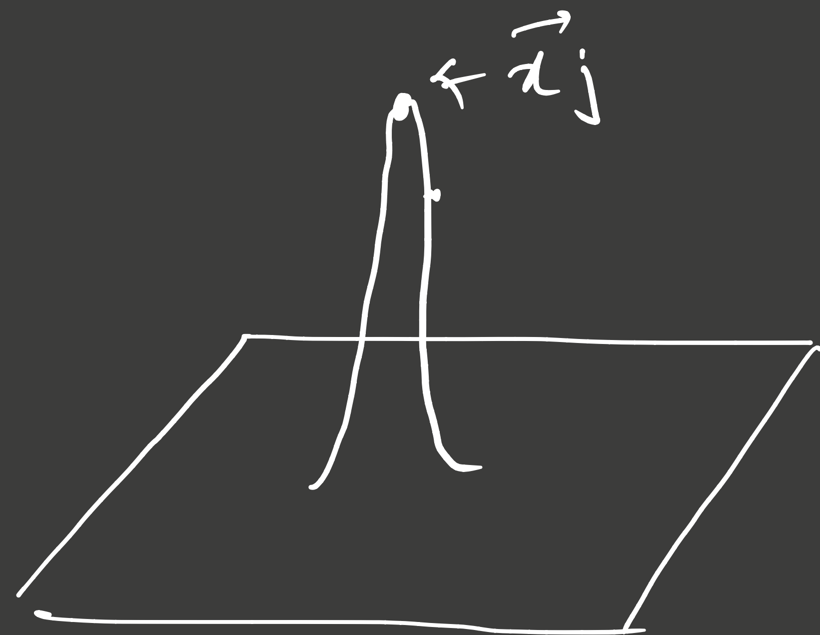
RBF : Effect of γ

γ: How far is the influence of a single training sample



γ = Low

High influence of $\bar{x}_j$

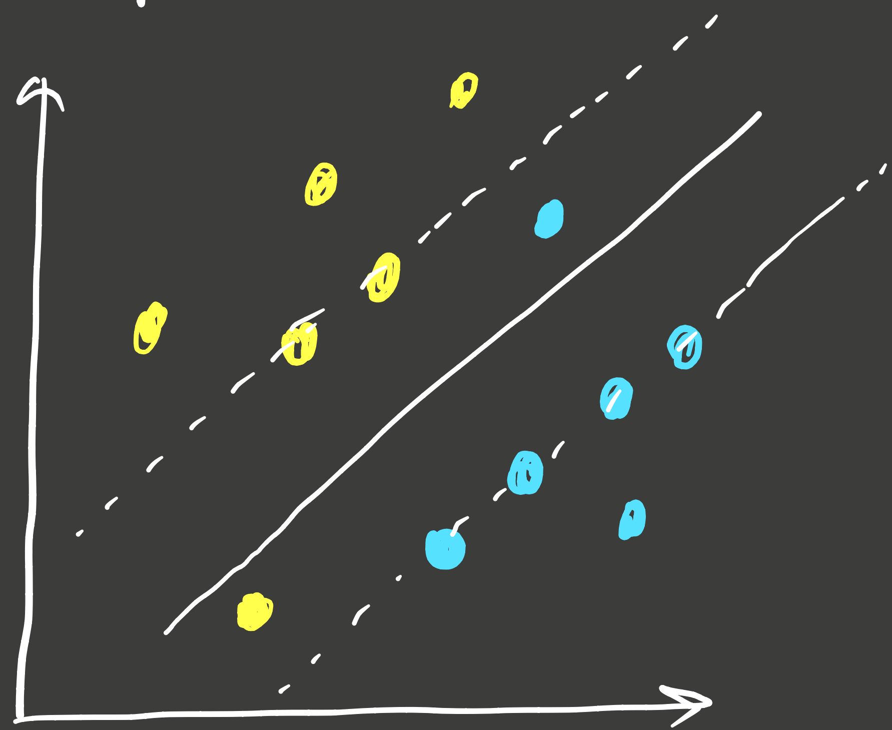


γ = High

low influence of $\bar{x}_j$

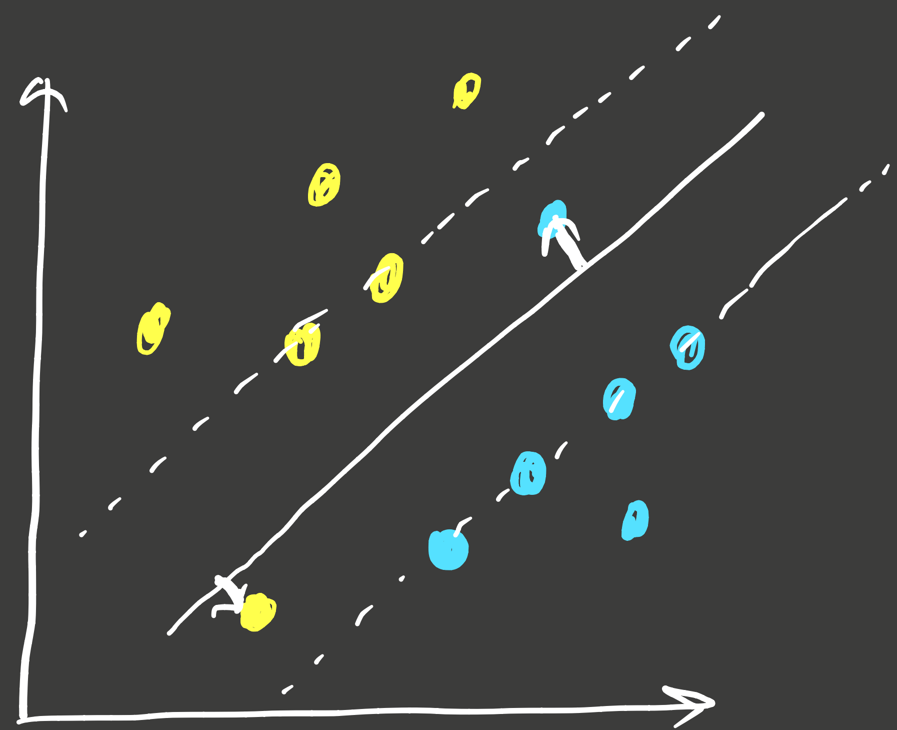
SOFT MARGIN SVM

Q:) Can we learn SVM for "slightly" non separable data without projecting to a higher space?



"Slightly" non separable data.

SOFT MARGIN SVM



SLACK VARIABLE

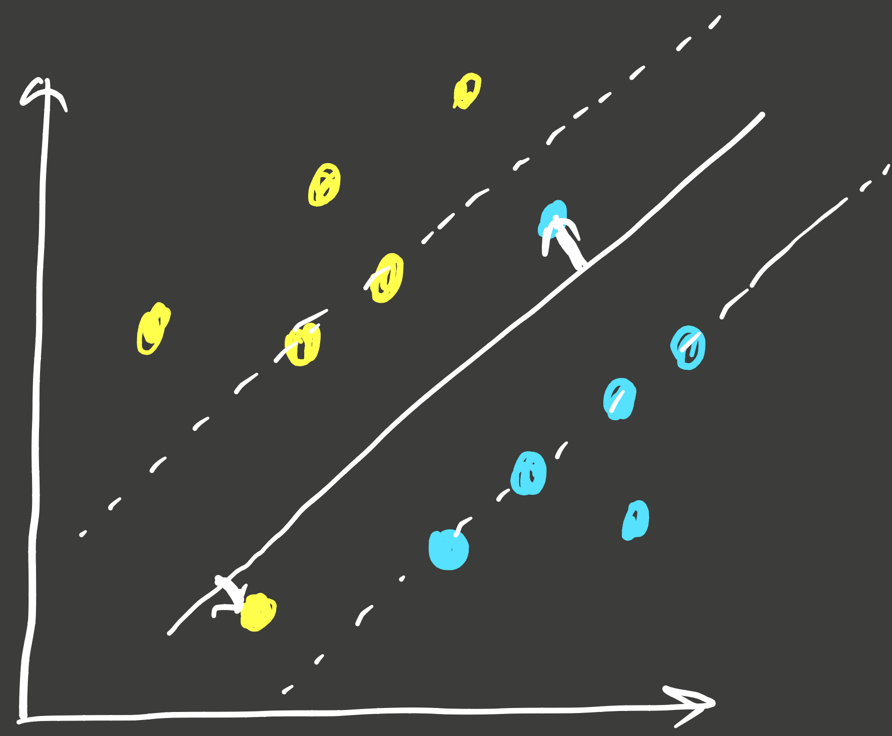
$$\xi_i = \begin{cases} 0 \text{ iff point on correct side} \\ \text{distance from hyperplane o/w} \end{cases}$$

Change objective

$$\min \frac{1}{2} \|\vec{w}\|^2 + C \sum_{i=1}^{n} \xi_i$$

s.t. $\quad y_i (\bar{w} \cdot \bar{x} + b) \geq 1 - \xi_i$

SOFT MARGIN SVM



Change objective

$$\min \frac{1}{2}\|\vec{w}\|^2 + C\sum_{i=1}^{n} \xi_i$$

s.t. $y_i(\bar{w}\cdot\bar{x}+b) \geq 1-\xi_i$

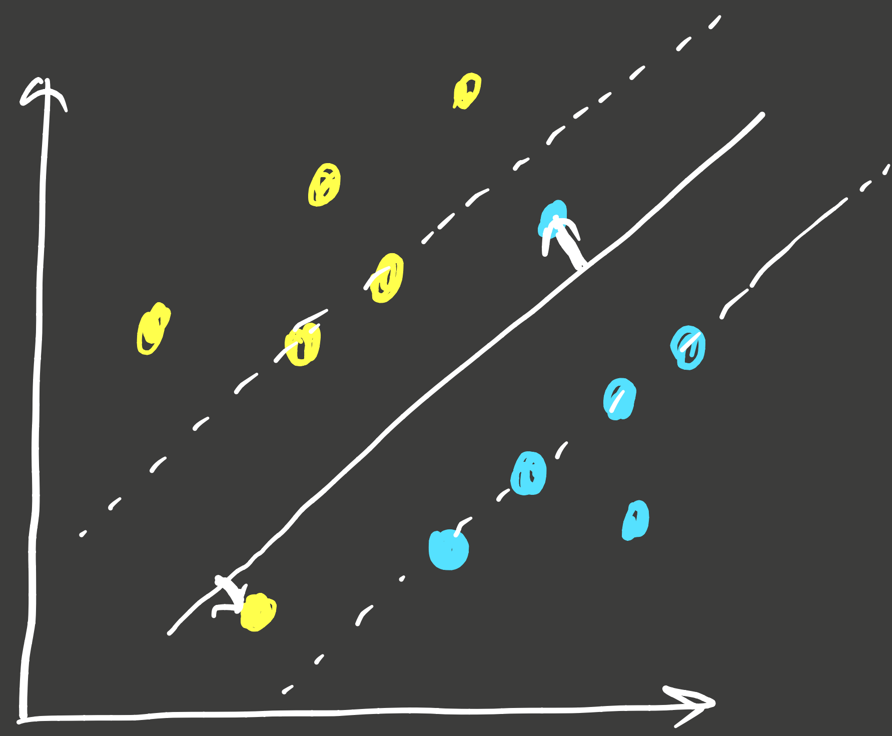
in Dual

Maximize $\sum_{i=1}^{n} \alpha_i - \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \bar{x}_i \cdot \bar{x}_j$

s.t.

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

SOFT MARGIN SVM



Change objective

$$\min \frac{1}{2}\|\vec{w}\|^2 + C\sum_{i=1}^{n} \xi_i$$

$$\text{s.t.} \quad y_i(\bar{w}\cdot\bar{x}+b) \geq 1-\xi_i$$

$C \to 0$: Larger margin

$C \to \infty$: smaller margin

SVM Loss + Penalty formulation
(Hinge loss)

SVC

SVR

why RBF is ∞ space

Taylor expansion

: has terms of $x, x^2 \ldots x^\infty$

Kernel function

$$K(x_i, x_j) = \phi(x_i)^T \cdot \phi(x_j)$$

$$\phi(x_i) = \begin{bmatrix} x_i \\ (x_i - 1)^2 \end{bmatrix} \qquad \phi(x_j) = \begin{bmatrix} x_j \\ (x_j - 1)^2 \end{bmatrix}$$

$$\phi(x_i)^T \cdot \phi(x_j) = \begin{bmatrix} x_i & (x_i - 1)^2 \end{bmatrix} \begin{bmatrix} x_j \\ (x_j - 1)^2 \end{bmatrix}$$

$$= x_i x_j + (x_i - 1)^2 (x_j - 1)^2$$

$$= x_i x_j + (x_i^2 + 1 - 2x_i)(x_j^2 + 1 - 2x_j)$$

$$= x_i x_j + (x_i^2 + 1 - 2x_i)(x_j^2 + 1 - 2x_j)$$

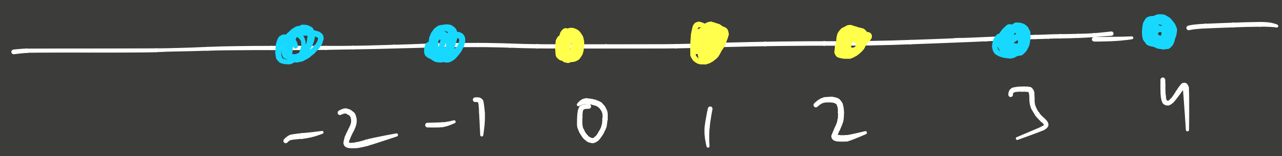
$$= x_i x_j + x_i^2 x_j^2 + x_i^2 - 2x_i^2 x_j$$
$$+ x_j^2 + 1 - 2x_j$$
$$- 2x_i x_j^2 - 2x_i + 4x_i x_j$$

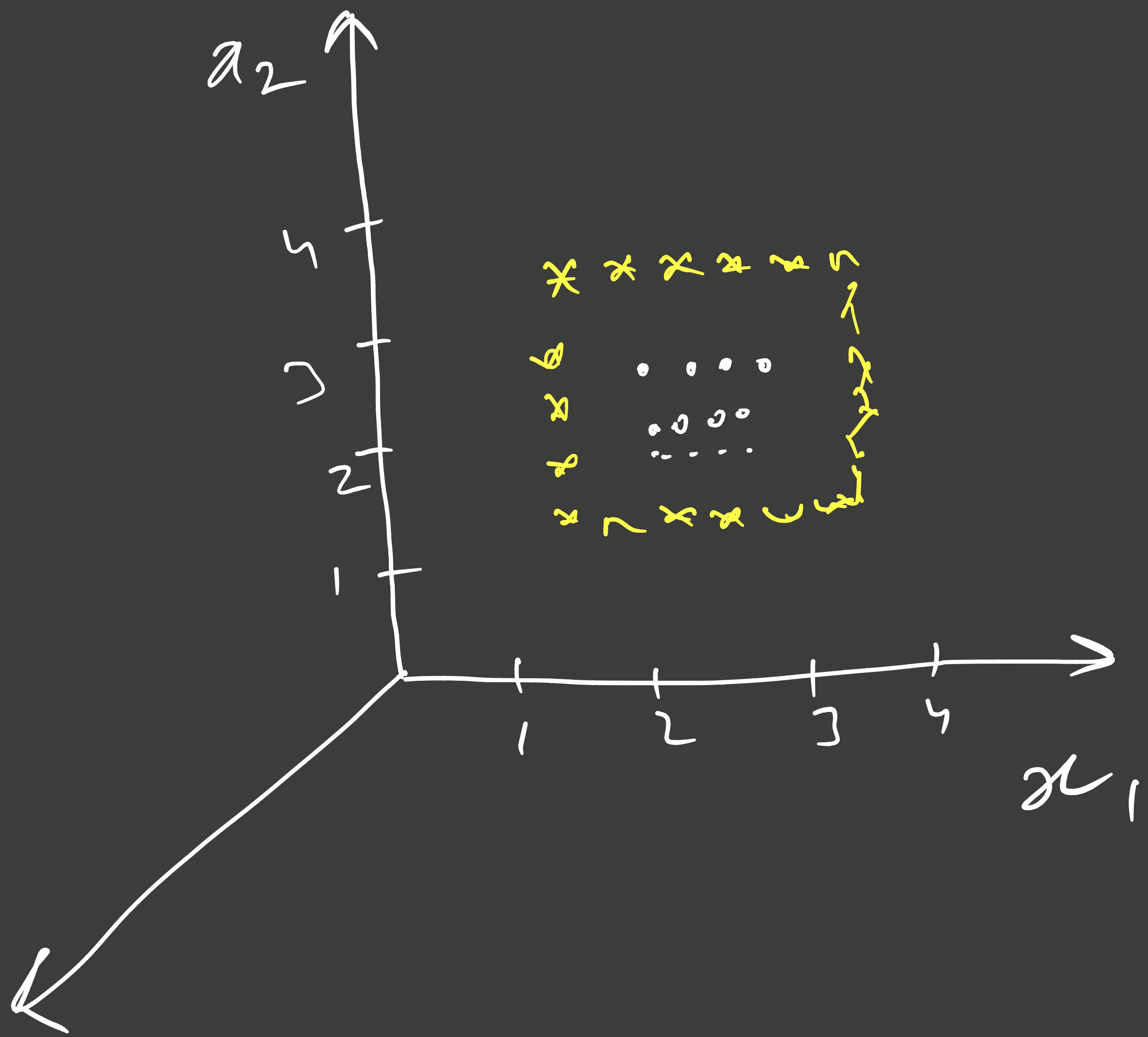
$$= 1 + 5x_i x_j + x_i^2 x_j^2 - 2x_i - 2x_j - 2x_i^2 x_j$$
$$- 2x_i x_j^2$$
$$+ x_i^2$$
$$+ x_j^2$$

$$= \langle 1, \sqrt{5} x_i, x_i^2,$$

$$K(x_i, x_j) = (1 + x_i^T x_j)^2$$

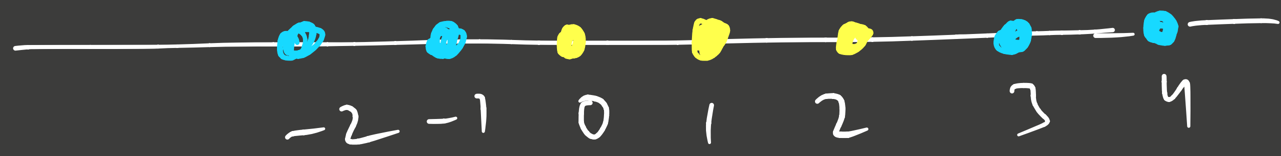
$$= 1 + 2x_i x_j + (x_i x_j)^2$$





$\sqrt{2}\,x\ldots$

$$(1+x_i x_j)^2 = 1+(\sqrt{2}x_i)(\sqrt{2}x_j)+x_i^2 x_j^2$$



$$\phi(x_i) = \langle 1, \sqrt{2}x_i, x_i^2 \rangle$$

$$\phi(x_j) = \langle 1, \sqrt{2}x_j, x_j^2 \rangle$$