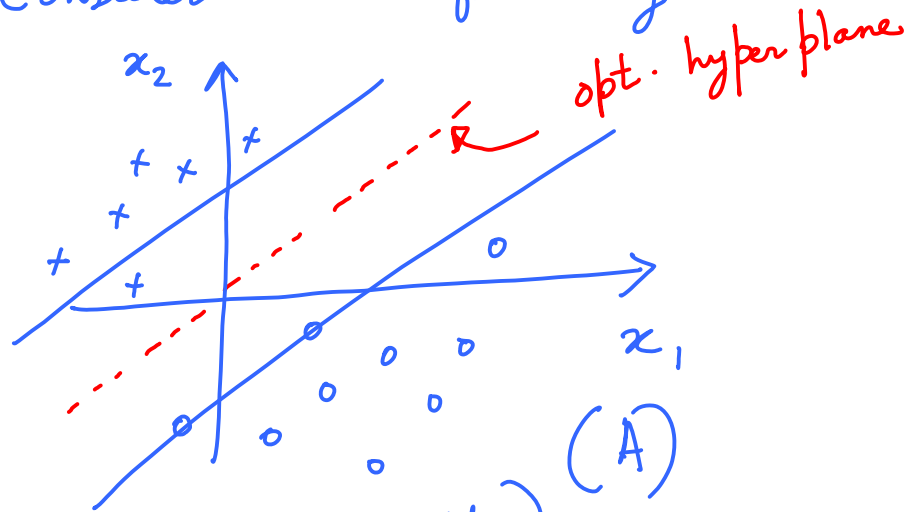


# Optimal hyperplane for non separable patterns

Consider the following cases



(Linearly Separable)

The points are on the correct side of the decision boundary



Some of the points are inside the region of separation but on the incorrect side

Let us introduce a new set of non negative scalar variables  $\{\xi_i\}_{i=1}^N$  into the defn. hyperplane

$$d_i (w^T x_i + b) \geq 1 - \xi_i \quad i = 1, \dots, N \quad \textcircled{I}$$

Slack variables

( Gives us a measure of deviation from ideal conditions of pattern separability )

For  $0 \leq \xi_i \leq 1$  Case (A)  
 For  $\xi_i > 1$ , it corresponds to Case (B)

The support vectors are those that satisfy the equality (precisely) even if  $\sum_i \xi_i > 0$  in  $\textcircled{I}$

GOAL: Find a separating hyperplane for which the error is minimized when averaged over the training set

Formulate 
$$\phi(\xi) = \sum_{i=1}^N I(\xi_i - 1)$$

$$I(\xi) = \begin{cases} 0 & \text{if } \xi \leq 0 \\ 1 & \text{else} \end{cases}$$

Indication function

To make the problem tractable,

$$\phi(\underline{z}) = \sum_{i=1}^N z_i$$

$$\phi(\underline{w}, \underline{z}) = \frac{1}{2} \underline{w}^T \underline{w} + C \sum_{i=1}^N z_i$$

$$\underline{z} = (z_1 \dots z_N)$$

objective function

Constraint

trade off in the complexity of the machine & the # of non-separable patterns

GOAL:

We need to optimize

$\phi(\underline{w}, \underline{z})$  w.r.t  $\underline{w}$  and  $\{z_i\}_{i=1}^N$

Let us formulate the primal and dual problems

---

PRIMAL : Given the training set  $\{\underline{x}_i, d_i\}_{i=1}^N$ ,  
find the opt. values of  $\underline{w}$  and  $b$  /  
$$d_i (\underline{w}^T \underline{x}_i + b) \geq 1 - \xi_i, \quad \forall i = 1, \dots, N$$
  
$$\xi_i \geq 0 \quad \forall i$$

We need to choose the wt. vector  $\underline{w}$  and slack variable  
 $\{\xi_i\}_{i=1}^N$  that minimize the cost functional  
$$\phi(\underline{w}, \underline{\xi}) = \frac{1}{2} \underline{w}^T \underline{w} + C \sum_{i=1}^N \xi_i$$
  
↑ user specified param.

Dual Problem : Find Lagrange multipliers  $\{\alpha_i\}_{i=1}^N$

---

that maximizes

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i d_i \alpha_j d_j x_i^T x_j$$

Subject to

$$(1) \quad \sum_{i=1}^N \alpha_i d_i = 0$$

$$(2) \quad 0 \leq \alpha_i \leq C \\ i = 1, \dots, N$$

↑ Observe the change:  
In case of linear separability  
 $\alpha_i \geq 0$

Note that neither slack variables  $\xi_i$  nor the Lagrange multipliers appear in the dual problem.

After proceeding with the opt. steps  $N_s \leftarrow \# \text{ support vectors}$

$$\underline{w}_{\text{opt}} = \sum_{i=1} \alpha_{\text{opt}, i} d_i \underline{x}_i$$

$$\alpha_i \left[ d_i \left( \underline{w}^T \underline{x}_i + b \right) - \left( 1 - \xi_i \right) \right] = 0 \quad i = 1, \dots, N$$

$$\mu_i \xi_i = 0 \quad ; \quad i = 1, \dots, N$$

Lagrange multipliers  $\mu_i$  to ensure that slack variables  $\xi_i$  are non negative

At the saddle point, for the primal problem

$$\frac{\partial J(\cdot)}{\partial \xi_i} = 0$$

$$\Rightarrow \alpha_i + \mu_i = C$$
$$\therefore \sum_i \mu_i = 0 \quad \text{if } \alpha_i < C$$

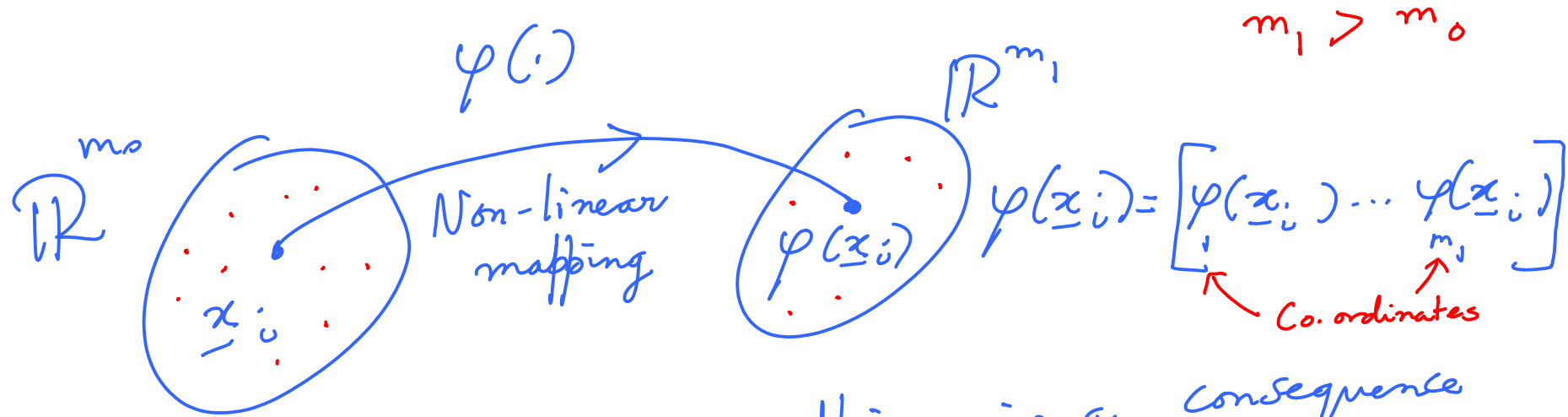


# Building SVMs for pattern recognition

---

Idea behind SVMs hinges on

- (1) Non-linear mapping of an input vector into a higher dimensional feature space hidden from both i/p and o/p.
- (2) Construction of an optimal hyperplane in lieu of (1) above.



The notion of non-linear mapping is a consequence of Cover's theorem that guarantees linear separability of patterns with high probability when input patterns are non-linearly transformed into a higher dim. feature space

# Inner Product Kernel

---

Let  $\underline{x}$  denote a vector drawn from an input space, say of dim.  $m_0$ .

Let  $\{\varphi_j(\underline{x})\}_{j=1}^{m_1}$  denote the set of non-linear transformations from the input to the feature space  $m_1$ : dim. of feature space

$\varphi_j(\underline{x})$  is defined a priori  $\forall j$   
 $\varphi(\underline{x}) = [\varphi_1(\underline{x}) \dots \varphi_j(\underline{x}) \dots \varphi_{m_1}(\underline{x})]^T$

Given a set of non-linear transformations,  
we can construct a hyperplane acting as  
a decision surface as follows:

$$\sum_{j=1}^{m_1} w_j \varphi_j(x) + b = 0$$

↖ Bias

where  $\{w_j\}_{j=1}^{m_1}$  denotes the set of linear  
weights connecting the feature space to the opp  
space along with bias 'b'.

# Folding the bias

$$\sum_{j=0}^{m_1} w_j \varphi_j(\underline{x}) = 0$$

$$\underline{w}^T \varphi = 0$$

$$j=0$$

$$w_0 = b$$

;

$$\varphi_0(\underline{x}) = 1$$

so that

$$\text{Let } \varphi(\underline{x}) = \begin{bmatrix} \varphi_0(\underline{x}) & \varphi_1(\underline{x}) & \dots & \varphi_{m_1}(\underline{x}) \end{bmatrix}^T$$

$$\varphi_0(\underline{x}) = 1$$

$$\forall \underline{x}$$

vector in  $\mathbb{R}^{m_1+1}$

In effect, the vector  $\varphi(\underline{x})$  is the 'image' of  $\underline{x}$  induced in the feature space.

$$\underline{w}^T \varphi(\underline{x}) = 0 \quad \text{-----} \quad (1)$$

Following w.r.t. Condition 1 in the derivative of Lagrangian with the transformation  $\varphi(\cdot)$  on  $\underline{x}$

$$\underline{w} = \sum_{i=1}^N \alpha_i d_i \varphi(\underline{x}_i) \quad \text{-----} \quad (2)$$

↑  
Notice the detail

$$\therefore \left[ \sum_{i=1}^N \alpha_i d_i \varphi^T(\underline{x}_i) \right] \varphi(\underline{x}) = 0$$

(Using ① in ②)

$\Downarrow \langle \varphi(\underline{x}_i), \varphi(\underline{x}) \rangle$  2 (4)

Now  $\varphi^T(\underline{x}_i) \varphi(\underline{x})$  represents the inner product of 2 vectors induced in the feature space by i/p  $\underline{x}$  and pattern  $\underline{x}_i$  in the training set.

Let us define the kernel  $K(\cdot)$

$$K(\underline{x}, \underline{x}_i) \triangleq \varphi^T(\underline{x}) \varphi(\underline{x}_i)$$

$$= \sum_{j=0}^{m_1} \varphi_j(\underline{x}) \varphi_j(\underline{x}_i)$$

*Coordinates of  $\varphi(\cdot)$*   
 $\forall i = 1, \dots, N$

This 'kernel' is symmetric in the arguments  
i.e.,  $K(\underline{x}, \underline{x}_i) = K(\underline{x}_i, \underline{x}) \quad \forall i$



In terms of the kernel, the  
opt. hyper plane is now given by

$$\sum_{i=1}^N \alpha_i d_i K(x, x_i) = 0 \quad (3)$$

Kernel

# MERCER'S THEOREM

(Mercer, 1908)

Let  $K(x, x')$  be a continuous symmetric kernel defined over the interval  $a \leq x \leq b$  and  $a \leq x' \leq b$ .

The kernel  $K(\cdot, \cdot)$  can be expanded as

$$K(x, x') = \sum_{i=0}^{\infty} \lambda_i \varphi_i(x) \varphi_i(x')$$

$\lambda_i > 0 \forall i$  (4)

For the expansion (4) to be valid,

and for it to be absolutely and

uniformly convergent,

$$\int_a^b \int_a^b K(x, x') \psi(x) \psi(x') dx dx' \geq 0$$

and

$$\int_a^b \psi^2(x) dx < \infty$$

Here, one can think of

$$\psi(x) \stackrel{\Delta}{=} \varphi^T(x) \varphi(x) = \langle \varphi(x), \varphi(x) \rangle$$

$$\text{where } \varphi(x) = \begin{bmatrix} \varphi_0(x) & \dots & \varphi_\infty(x) \end{bmatrix}^T$$

Example:  $\psi(x) = (1+x)^2$  over  $[0, 1]$

$$\psi(x) = 1 + 2x + x^2 = \begin{pmatrix} 1 & \sqrt{2}x & x \end{pmatrix} \begin{pmatrix} 1 & \sqrt{2}x & x \end{pmatrix}^T$$

The functions  $\{\psi_i(x)\}$  are called the  
eigen functions of the expansion and

$\lambda_i$ s are the eigen values

$\lambda_i > 0 \implies K(x, x')$  is +ve definite

Let us explore this for a vector  $\underline{x}$  i.e. a  
 $(1+x)^2$ ;  $(1+\underline{x}^T \underline{x})^2$   $\underline{x} \in \mathbb{R}^d$

In light of Mercer's theorem,

For  $\lambda_i \neq 1$ , the  $i^{\text{th}}$  image

$\sqrt{\lambda_i} \varphi_i(\underline{x})$  induced in the feature space

by the  $i^{\text{th}}$  coord. in  $\varphi(\underline{x})$  is an eigen function  
of the expansion.

2) In theory, the dim. of feature  
space can be  $\infty$ !

Mercer's Theorem tells us if a candidate  
kernel is actually an inner product kernel in  
some space  $\Rightarrow$  admissible within the  
SVM design framework

# Examples of I.P. kernels

---

Machine	I.P. Kernel	Remarks
Polynomial learning machine	$(x^T x_i + 1)^p$	Power 'p' specified a priori
Radial basis function	$\exp\left(-\frac{1}{2\sigma^2} \ x - x_i\ ^2\right)$	Width $\sigma^2$ common to all kernels, specified a priori or <u>estimated adaptively</u>



# Optimum Design of a SVM

---

The expansion of the inner product kernel  $K(\underline{x}, \underline{x}_i)$  allows us to construct a decision surface which is non-linear in the i/p space, but with image in the feature space 'linear'.

GOAL: Set up the optimization problem using  $K(\cdot)$

$$\underline{\alpha} = (\alpha_1, \dots, \alpha_N)$$

Construct the optimization problem

$$\max_{\underline{\alpha}} Q(\underline{\alpha}) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j K(\underline{x}_i, \underline{x}_j)$$

Subject to

$$(1) \quad \sum_{i=1}^N \alpha_i d_i = 0$$

$$(2) \quad 0 \leq \alpha_i \leq C$$

$$i = 1, \dots, N$$

$$b = w_0 \text{ for } \varphi_0(\underline{x}) = 1$$

Observe  
the  $K(\cdot, \cdot)$   
as against  
 $\underline{x}_i^T \underline{x}_j$  in  
linearly separable  
situation in the  
i/p space.

The inner products  $\underline{x}_i^T \underline{x}_j$  we dealt with under SVM design for linear separability

is replaced with  $K(\underline{x}_i, \underline{x}_j)$

Define a symmetric matrix

$$K := \left[ K(\underline{x}_i, \underline{x}_j) \right]_{i,j=1}^N$$

$$\begin{bmatrix} K(\underline{x}_1, \underline{x}_1) & \dots & K(\underline{x}_1, \underline{x}_N) \\ \vdots & & \vdots \\ K(\underline{x}_N, \underline{x}_1) & \dots & K(\underline{x}_N, \underline{x}_N) \end{bmatrix}$$

Gram matrix

After having found opt. values for the  
Lagrange multipliers  $\alpha_{opt, i}$  / compute

opt.  $\frac{w}{opt}$  as

$$\frac{w}{opt} = \sum_{i=1}^N \alpha_{opt, i} d_i \varphi(\underline{x}_i)$$

Image induced in the feature space

## $\epsilon$ -insensitive loss function

GOAL : Come up with a robust estimator for a regression problem insensitive to small changes in the model.

Problems seen : A least squares (L.S) estimator is sensitive to the presence of outliers.  $\epsilon$  performs poorly over distributions with long tails.

Idea

Minimize max. degradation.

Construct a loss function

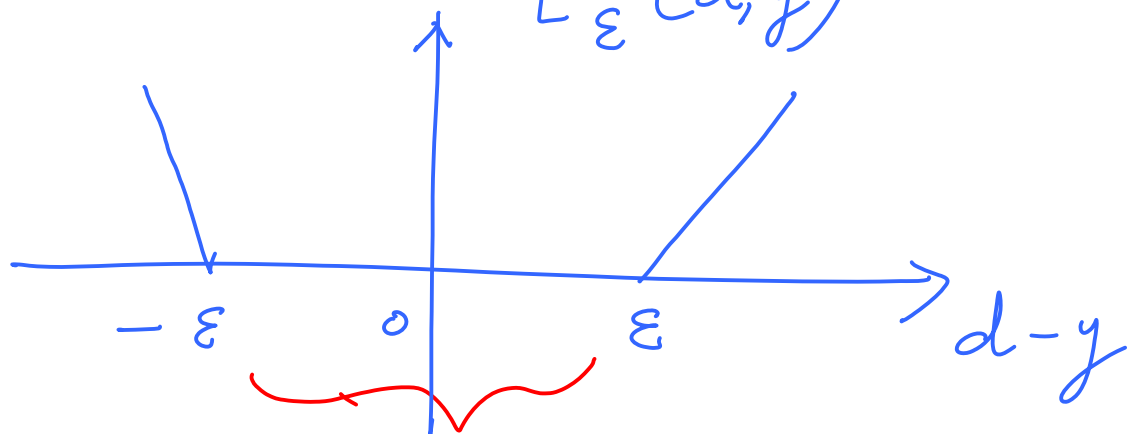
$$L(d, y) = |d - y|$$

← desired response  
← estimator  $\hat{y}$

Define  $L_{\varepsilon}(d, y) = \begin{cases} |d - y| - \varepsilon; & |d - y| \geq \varepsilon \\ 0; & \text{else} \end{cases}$

' $\epsilon$ ' is a prescribed parameter  
 $L_\epsilon(d, y)$

$\epsilon$ -insensitive  
loss function



insensitive up to an  
 $\epsilon$

# SVMs for non linear regression

---

Consider a non linear regression model

$$d = f(\underline{x}) + v$$

$f(\cdot)$  is unknown, Statistics of  $v$  is unknown

We have the training set  $\{ \underline{x}_i, d_i \}_{i=1}^N$

GOAL : Get an estimate of the dependence of  $d$  on  $\underline{x}$



Let 'd' be estimated by 'y' using a

linear combination of some non linear basis functions

$$\left\{ \varphi_j(\underline{x}) \right\}_{j=0}^{m_1}$$

$$\langle \underline{w}, \varphi(\underline{x}) \rangle$$

$$y = \sum_{j=0}^{m_1} w_j \varphi_j(\underline{x}) = \underline{w}^T \varphi(\underline{x})$$

where

$$\varphi(\underline{x}) = \begin{bmatrix} \varphi_0(\underline{x}) & \dots \\ \varphi_{m_1}(\underline{x}) \end{bmatrix}; \varphi_0(\underline{x}) = 1$$

$$\underline{w} = \begin{bmatrix} w_0 & \dots \\ w_{m_1} \end{bmatrix}^T; w_0 = b$$

Minimize the empirical risk

$$R_{emp} = \frac{1}{N} \sum_{i=1}^N L_{\varepsilon}(d_i, y_i)$$

Subject to  $\| \underline{w} \|^2 \leq C_0$ ;  $C_0$  is a constant

Let us formulate it as a constrained opt. problem by introducing 2 sets of slack variables

$$\left\{ \xi_i \right\}_{i=1}^N \quad \text{and} \quad \left\{ \xi'_i \right\}_{i=1}^N$$

$$\left\{ \begin{array}{l} d_i - \underline{w}^T \varphi(x_i) \leq \varepsilon + \xi_i; \quad i=1, \dots, N \\ \underline{w}^T \varphi(x_i) - d_i \leq \varepsilon + \xi'_i; \quad i=1, \dots, N \\ \xi_i \geq 0 \\ \xi'_i \geq 0 \end{array} \right.$$



The constrained opt. problem is equivalent to minimizing the cost functional

$$\phi(\underline{w}, \underline{\xi}, \underline{\xi}') = C \sum_{i=1}^N (\underline{\xi}_i + \underline{\xi}'_i) + \frac{1}{2} \underline{w}^T \underline{w}$$

misclassified points norm<sup>2</sup> of the  $\underline{w}$

Forming the Lagrangian

$$J(\underline{w}, \underline{\xi}, \underline{\xi}', \underline{\alpha}, \underline{\alpha}', \underline{\gamma}, \underline{\gamma}')$$

in terms of the

Cost as the Lagrange

constraints using  $\underline{\alpha}, \underline{\alpha}', \underline{\gamma}, \underline{\gamma}'$  multipliers

$$\begin{aligned}
 J(\cdot) &= C \sum_{i=1}^N (\xi_i + \xi_i') + \frac{1}{2} \underline{\omega}^T \underline{\omega} \\
 &\quad - \sum_{i=1}^N \alpha_i (\underline{\omega}^T \varphi(\underline{x}_i) - d_i + \varepsilon + \xi_i) \\
 &\quad - \sum_{i=1}^N \alpha_i' (d_i - \underline{\omega}^T \varphi(\underline{x}_i) + \varepsilon + \xi_i') \\
 &\quad - \sum_{i=1}^N \gamma_i \xi_i - \sum_{i=1}^N \gamma_i' \xi_i'
 \end{aligned}$$

Cost

} Conditions for the constraints over  $L_\varepsilon(\cdot)$

} Conditions for slack variables

Taking  $\frac{\partial J(\cdot)}{\partial \underline{\omega T}} = 0$

$\Rightarrow \underline{\omega} = \sum_{i=1}^N (\alpha_i - \alpha_i') \varphi(\underline{x}_i)$

Verify :  $\left. \begin{matrix} \alpha_i \\ \alpha_i' \end{matrix} \right|_{i=1}^N = \left. \begin{matrix} C - \alpha_i \\ C - \alpha_i' \end{matrix} \right\}$  Putting a constraint on  $\alpha_i, \alpha_i'$

One can formulate the dual problem towards maximization

$$Q(\underline{\alpha}, \underline{\alpha}') = \sum_{i=1}^N d_i (\alpha_i - \alpha_i') - \sum_{i=1}^N (\alpha_i + \alpha_i')$$

Inner product  
kernel  
↓

$$\text{where } K(\cdot, \cdot) = \frac{1}{2} \varphi^T(\underline{x}_i) \varphi(\underline{x}_j) \text{ is an I.P. kernel } \langle \varphi(\underline{x}_i), \varphi(\underline{x}_j) \rangle$$

NOTE:

Given  $\{x_i, d_i\}_{i=1}^N$ , find  $\{\alpha_i\}_{i=1}^N$   
and  $\{\alpha_i'\}_{i=1}^N$  that max.  $Q(\alpha_i, \alpha_i')$   
Subject to (1)  $\sum_{i=1}^N \alpha_i - \alpha_i' = 0$   
(2)  $0 \leq \alpha_i, \alpha_i' \leq C$   
 $i=1, \dots, N$

Training Set



Now, with  $\varepsilon$  and  $C$  as our  
free parameters

$$F(\underline{x}, \underline{w}) = \frac{w^T \varphi(\underline{x})}{\sum_{i=1}^N (\alpha_i - \alpha_i')} K(\underline{x}, \underline{x}_i)$$

# XOR Problem Revisited

---

Let us consider the XOR problem using SVMs  
(Cherkassy, 1998)

Start with a kernel

$$K(\underline{x}, \underline{x}_i) = \left( 1 + \underline{x}^T \underline{x}_i \right)^2$$

$$\underline{x} = [x_1 \ x_2]^T$$

$$\underline{x}_i = [x_{i1} \ x_{i2}]^T$$

$$\begin{aligned}
k(\underline{x}, \underline{x}_i) &= \left( 1 + (\underline{x}_1 \ \underline{x}_2) \cdot \begin{bmatrix} x_{i1} \\ x_{i2} \end{bmatrix} \right)^2 \\
&= \left( 1 + x_1 x_{i1} + x_2 x_{i2} \right)^2 \\
&= 1 + x_1^2 x_{i1}^2 + 2 x_1 x_2 x_{i1} x_{i2} \\
&\quad + x_2^2 x_{i2}^2 + 2 x_1 x_{i1} + 2 x_2 x_{i2}
\end{aligned}$$

Let us express  $k(\cdot, \cdot) = \langle \varphi(\underline{x}), \varphi(\underline{x}_i) \rangle^T$

$$\begin{aligned}
\varphi(\underline{x}) &= \begin{bmatrix} 1 & x_1^2 & \sqrt{2} x_1 x_2 & x_2^2 & \sqrt{2} x_1 & \sqrt{2} x_2 \end{bmatrix}^T \\
\varphi(\underline{x}_i) &= \begin{bmatrix} 1 & x_{i1}^2 & \sqrt{2} x_{i1} x_{i2} & x_{i2}^2 & \sqrt{2} x_{i1} & \sqrt{2} x_{i2} \end{bmatrix}^T
\end{aligned}$$

For the XOR problem,

$(-1, -1) \rightarrow -1$   
 $(-1, +1) \rightarrow +1$   
 $(+1, -1) \rightarrow +1$   
 $(+1, +1) \rightarrow -1$

$$K := [K(\underline{x}_i, \underline{x}_j)]$$

$$K(\underline{x}_i, \underline{x}_j) = \varphi^T(\underline{x}_i) \varphi(\underline{x}_j)$$

Each  $\underline{x}_i, \underline{x}_j$


$$K = \begin{bmatrix} 9 & 1 & 1 & 1 \\ 1 & 9 & 1 & 1 \\ 1 & 1 & 9 & 1 \\ 1 & 1 & 1 & 9 \end{bmatrix}_{4 \times 4}$$

From the dual problem, the objective  $Q(\underline{\alpha})$

$$Q(\underline{\alpha}) = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 - \frac{1}{2} \left( \begin{aligned} &9\alpha_1^2 - 2\alpha_1\alpha_2 - 2\alpha_1\alpha_3 \\ &\quad + 2\alpha_1\alpha_4 \\ &\quad + 9\alpha_2^2 + 2\alpha_2\alpha_3 - 2\alpha_2\alpha_4 \\ &\quad + 9\alpha_3^2 - 2\alpha_3\alpha_4 + 9\alpha_4^2 \end{aligned} \right)$$

$\sum_{i=1}^4 \sum_{j=1}^4 \alpha_i \alpha_j d_i d_j K(\underline{x}_i, \underline{x}_j)$

$$\frac{\partial Q(\underline{\alpha})}{\partial \alpha_i} = 0 \quad i = 1, \dots, 4$$

Doing this yields

$$\left\{ \begin{array}{l} 9\alpha_1 - \alpha_2 - \alpha_3 + \alpha_4 = 1 \\ -\alpha_1 + 9\alpha_2 + \alpha_3 - \alpha_4 = 1 \\ -\alpha_1 + \alpha_2 + 9\alpha_3 - \alpha_4 = 1 \\ \alpha_1 - \alpha_2 - \alpha_3 + 9\alpha_4 = 1 \end{array} \right.$$

Solving this set  
of eqns

$$\alpha_{\text{opt}, i} = \frac{1}{8} \\ i = 1, \dots, 4$$

Now, all 4 inputs  $\{\underline{x}_i\}_{i=1}^4$  are support vectors

$$Q(\alpha) = \frac{1}{4}$$

$$\frac{1}{2} \|\underline{w}_0\|^2 = \frac{1}{4} \Rightarrow \|\underline{w}_0\| = \frac{1}{\sqrt{2}}$$

$$\underline{w}_0 = \sum_{i=1}^4 \alpha_{0,i} d_i \varphi(\underline{x}_i)$$
$$= \frac{1}{8} \left[ -\varphi(\underline{x}_1) + \varphi(\underline{x}_2) + \varphi(\underline{x}_3) - \varphi(\underline{x}_4) \right]$$

$$\underline{w}_0 = \begin{bmatrix} 0 & 0 & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 \end{bmatrix}^T$$

The opt. hyperplane is given by Linear framework

$$\underline{w}_0^T \varphi(\underline{x}) = 0$$

$$\begin{bmatrix} 0 & 0 & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ x_1^2 \\ \sqrt{2} x_1 x_2 \\ x_2^2 \\ \sqrt{2} x_1 \\ \sqrt{2} x_2 \end{bmatrix} = 0$$

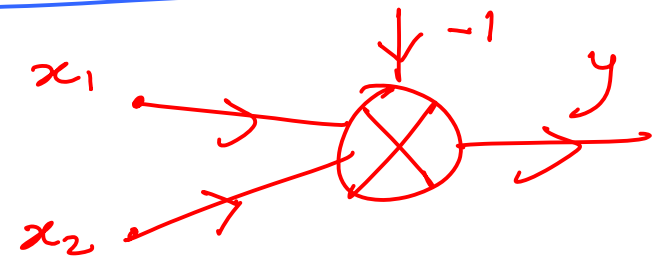
$\Rightarrow x_1 x_2 = 0$  is the decision boundary



It is nonlinear

$\underline{y}$	$= -x_1 x_2$
$\underline{x}$	$y = -x_1 x_2$
$(-1, -1)$	$-1$
$(-1, +1)$	$+1$
$(+1, -1)$	$+1$
$(+1, +1)$	$-1$

as desired!



## Hilbert Space

Let  $\{\underline{x}_k\}_{k=1}^{\infty}$  be an orthonormal basis for an inner product space  $\mathcal{F}$  (possibly of  $\infty$  dimensions)

$$\langle \underline{x}_j, \underline{x}_k \rangle = \underline{x}_j^T \underline{x}_k = \begin{cases} 1 & j = k \\ 0 & \text{else} \end{cases}$$

Let  $\mathcal{H}$  be the largest and most inclusive space of vectors for which the set  $\{\underline{x}_k\}_{k=1}^{\infty}$  is a basis.

Then any vector  $\underline{x}$  not necessarily lying in  $\mathcal{F}$  can be written as  $\underline{x} = \sum_{k=1}^{\infty} a_k \underline{x}_k$ .  $\underline{x}^* = \lim_{n \rightarrow \infty} \underline{x}^{(n)} \notin \mathcal{F}$   
Converging vector (subtle issue)

Define a new vector

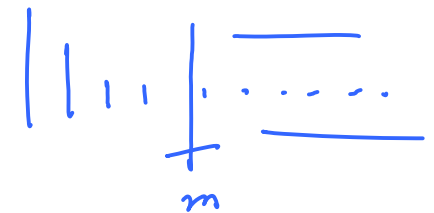
$$\underline{y}_n = \sum_{k=1}^n a_k \underline{x}_k$$

$$\text{Similarly } \underline{y}_m = \sum_{k=1}^m a_k \underline{x}_k$$

For  $n > m$

$$\begin{aligned} \|\underline{y}_n - \underline{y}_m\|^2 &= \left\| \sum_{k=1}^n a_k \underline{x}_k - \sum_{k=1}^m a_k \underline{x}_k \right\|^2 \\ &= \left\| \sum_{k=m+1}^n a_k \underline{x}_k \right\|^2 \\ &< \sum_{k=m+1}^n a_k^2 \underbrace{\|\underline{x}_k\|^2}_{\text{norm. 1}} = \sum_{k=m+1}^n a_k^2 \end{aligned}$$

$$\begin{array}{l}
 1) \quad \sum_{k=m+1}^n a_k^2 \xrightarrow{m, n \rightarrow \infty} 0 \\
 2) \quad \sum_{k=1}^n a_k^2 < \infty
 \end{array}$$



Pick an  $\epsilon > 0$  and  $m \gg \frac{1}{\epsilon}$  /  $\sum_{k=m+1}^{\infty} a_k^2 < \epsilon$  ①

Since  $\sum_{k=1}^{\infty} a_k^2 = \sum_{k=1}^m a_k^2 + \sum_{k=m+1}^{\infty} a_k^2$

$$\sum_{k=1}^{\infty} a_k^2 < \infty$$

A sequence of vectors in a normed space, for which the Euclidean distance  $\|y_n - y_m\| < \varepsilon$  for any  $\varepsilon > 0$  and  $m, n > M$  is a convergent sequence called a Cauchy sequence.

$\{y_n\}_{n=1}^{\infty}$  is Cauchy

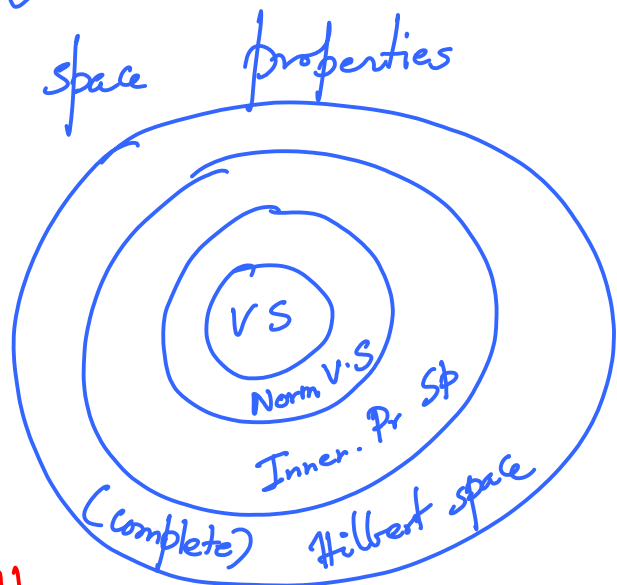
$$\|y_n - y_m\| \xrightarrow{m, n \rightarrow \infty} 0$$

- (Hilbert space)  $\mathcal{H}$  is more complete than  $\mathcal{F}$  (I.P)
- 1) Every Cauchy sequence of vectors taken from  $\mathcal{H}$  converges to a limit in  $\mathcal{H}$ .
  - 2)  $\mathcal{H}$  inherits the inner product space properties

NOTE: Are all Cauchy sequences convergent? No

$$\sum_{k \geq 0} \frac{1}{k!} \rightarrow e \notin \mathbb{Q}$$

over all rational nos  $\mathbb{Q}$



All convergent sequences are Cauchy  
 But, all Cauchy sequences are not convergent