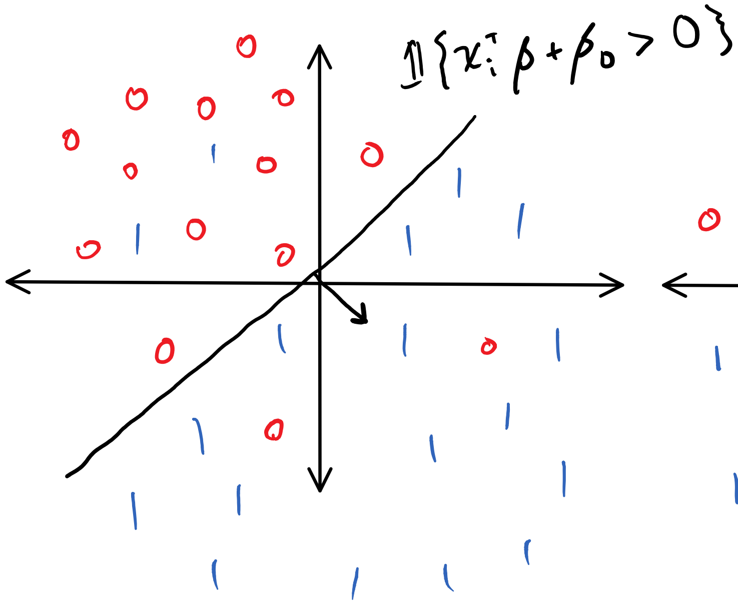
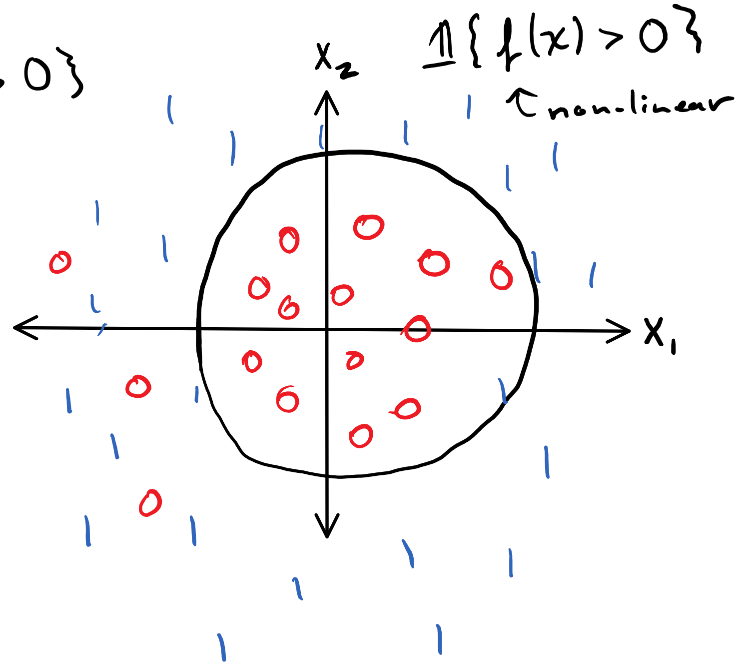
HiDi Embedding

Wednesday, May 8, 2019 1:20 PM



Linear Decision Boundary



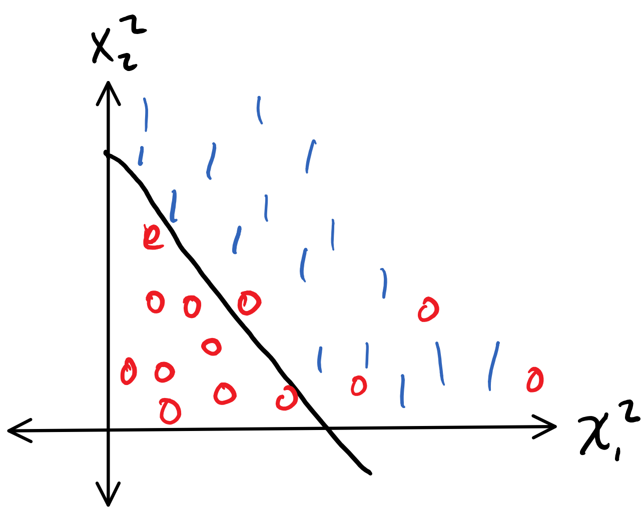
Non-linear decision boundary

<u>def</u> higher-dimensional embedding

$$\Phi : \mathbb{R}^p \to \mathbb{R}^D \qquad \Phi(x) \in \mathbb{R}^D$$

<u>ex</u> $\Phi(x_1, x_2)$

$$= (1, x_1, x_2, x_1^2, x_2^2)$$



Φ makes linear methods into non-linear ones

<u>ex</u> $x_1, \ldots, x_p$ are proposition and we want to

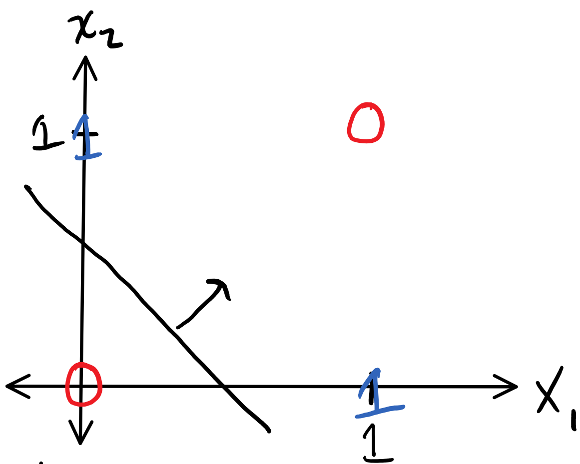
answer ~~logi~~ complex props such as $(x_1 \text{ or } x_2) \text{ and } (x_3 \text{ or } x_4)$

x_1 xor x_2 : $(x_1 \text{ and not } x_2) \text{ or } (x_2 \text{ and not } x_1)$

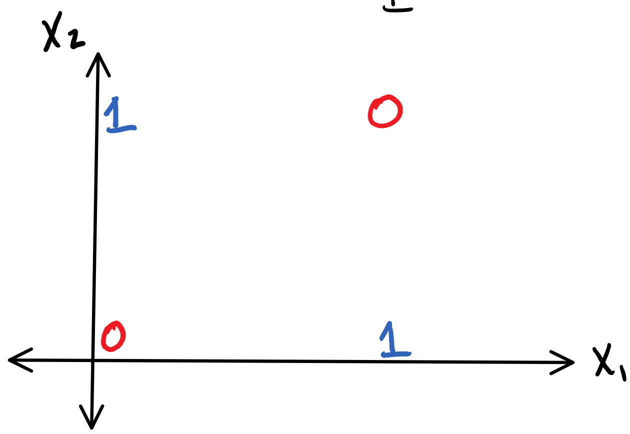
$x_1, x_2 \in \{0, 1\}$ False True

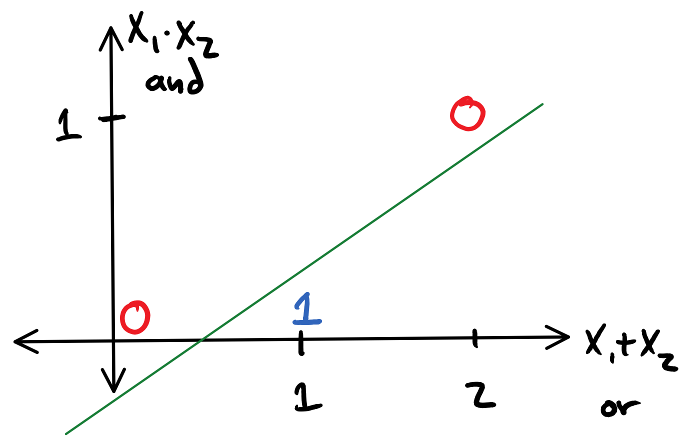
$y = x_1 \text{ xor } x_2$

y	x_1	x_2
0	0	0
1	0	1
1	1	0
0	1	1



$$\Phi(x_1, x_2) = (x_1, x_2, x_1 \cdot x_2)$$





Kernel Trick

Wednesday, May 8, 2019 1:23 PM

Let $Z_{ik} = \Phi_k(x_i) \quad x_i \in \mathbb{R}^p \quad k=1,\dots,D$

linear - SVM for $y_i \in \{-1, 1\}$

$$\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \overset{(I)}{\left(1 - y_i \underset{\text{row}}{z_i^T} \beta\right)_+} + \overset{(II)}{\lambda \|\beta\|^2}$$

<u>claim</u> $\hat\beta$ solves SVM can be written $Z^T\alpha$, $\alpha \in \mathbb{R}^n$

i.e. $\hat\beta_j = \underset{\text{column}}{Z_j^T} \alpha = \sum_i \alpha_i Z_{ij} = \left(\sum_i \alpha_i z_i\right)_j$

$\beta = \underbrace{\sum_i \alpha_i z_i}_{Z^T\alpha} + \beta^\perp \qquad z_i^T \beta^\perp = 0 \quad \forall i$

$z_i^T\beta = \sum_j \alpha_j z_i^T z_j + \cancelto{0}{z_i^T \beta^\perp}$

$= z_i^T Z^T \alpha \quad \Rightarrow \beta^\perp$ does not impact (I)

$\|\beta\|^2 = \|Z^T\alpha + \beta^\perp\|^2 = \|Z^T\alpha\|^2 + \underbrace{2\beta^{\perp T} Z^T \alpha}_{=\alpha^T \left(\cancelto{0}{Z\beta^\perp}\right)} + \|\beta^\perp\|^2$

$= \|Z^T\alpha\|^2 + \|\beta^\perp\|^2$

$\Rightarrow \beta^\perp$ at the min is 0

Rewrite SVM:

$$\min_{\alpha \in \mathbb{R}^n} \frac{1}{n} \sum_i \left(1 - y_i z_i^T Z^T \alpha\right)_+ + \lambda \|Z^T\alpha\|^2$$

$$\hat\beta = Z^T\hat\alpha = \sum_i \hat\alpha_i x_i$$

<u>General</u>

$\min \quad R(y, Z\beta) + \lambda \|\beta\|^2$

$$\min_{\beta \in \mathbb{R}^D} R_n(y, Z\beta) + \lambda \|\beta\|^2$$

$$|||$$

$$\min_{\alpha \in \mathbb{R}^n} R_n(y, K\alpha) + \lambda \alpha^T K \alpha$$

w/ $K = ZZ^T$ $\quad (z_i^T Z^T \alpha) = (K\alpha)_i$

↳ kernel matrix

$$K_{ij} = z_i^T z_j = \underline{\Phi}(x_i)^T \underline{\Phi}(x_j) \quad (1)$$

$$\hat{\beta} = Z^T \hat{\alpha}$$

Method 1: define $\underline{\Phi}$, apply $\to Z$, make K

Method 2: define <u>kernel function</u>

$k(x_i, x_j) = \underline{\Phi}(x_i)^T \underline{\Phi}(x_j)$ (closed form)

compute K from (1)

predict $\underline{\Phi}(x)^T \hat{\beta} = \underline{\Phi}(x)^T \left(\sum_i \hat{\alpha}_i z_i\right)$

$$= \sum_i \hat{\alpha}_i \underline{\Phi}(x)^T \underline{\Phi}(x_i)$$

$$= \sum_i \hat{\alpha}_i k(x, x_i)$$

<u>ex</u> d^{th} degree poly. $k(x, x') = (1 + x^T x')^d$

$d=2$: $(1 + x_1 x_1' + x_2 x_2')^2 = 1 + 2x_1 x_1' + 2x_2 x_2' + x_1^2 x_1'^2 + x_2^2 x_2'^2 + 2x_1 x_1' x_2 x_2'^2$

$$= \underbrace{(1, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1 x_2, x_1^2, x_2^2)}^T (\ldots x' \ldots)$$