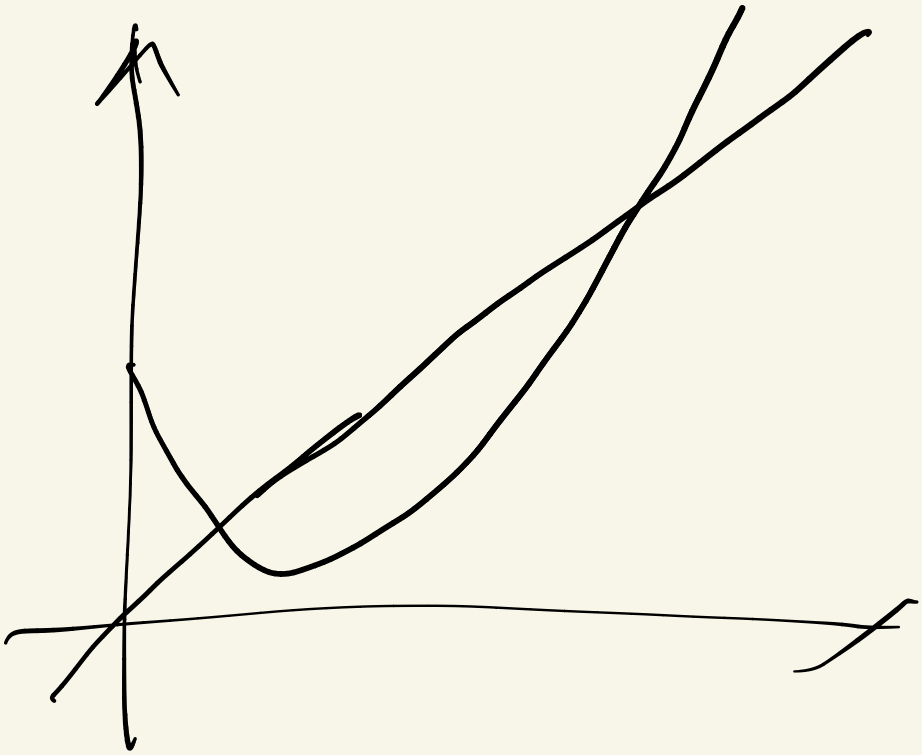
Lecture 5 Kernel & SVM



n : # data samples

$\phi(x) \in R^p \quad p = 1000000$

$\boxed{\theta^T \phi(x)}$

$$\boxed{\phi(x^{(j)})^T \phi(x^{(i)})} \quad \beta_j \in \mathbb{R}$$

$\phi(x) \in \mathbb{R}^p$

$\theta^T \phi(x)$

n sample

n × n

$K = \begin{bmatrix} & & \\ & & \end{bmatrix}$ n

n × n

m # gradient steps

$\phi(x)$

$\langle \phi(x^{(i)}), \phi(x^{(j)}) \rangle$

$K = \begin{bmatrix} \\ \\ \end{bmatrix}$

$$K(x,z) = \phi(x)^T \phi(z) \qquad (x^T z)^2$$

$$K(x,z) = (x^T z)^2 \quad O(d)$$

$$= \left(\sum_{i=1}^{d} x_i z_i\right)\left(\sum_{j=1}^{d} x_j z_j\right)$$

$$\left(\sum_i a_i\right)^2 = \sum_i \sum_j a_i a_j$$

$$= \sum_{i=1}^{d} \sum_{j=1}^{d} x_i x_j z_i z_j$$

$$= \sum_{i,j=1}^{d} (x_i x_j)(z_i z_j) \qquad d^2$$

explicit

$$\phi(x) \longrightarrow \phi(x)^T \phi(z)$$

implicit

$K(x, z)$

$\phi(x)$?

$$\underbrace{K(x,z) = (x^T z)^K}_{O(d)} \rightarrow \underbrace{O(d^K)}_{O(d^K)}$$

positive semi definite

Any non-zero vector $z \in R^d$, $A \in R^{d \times d}$

$z^T A z \geq 0$ full rank, invertible

$\Updownarrow$

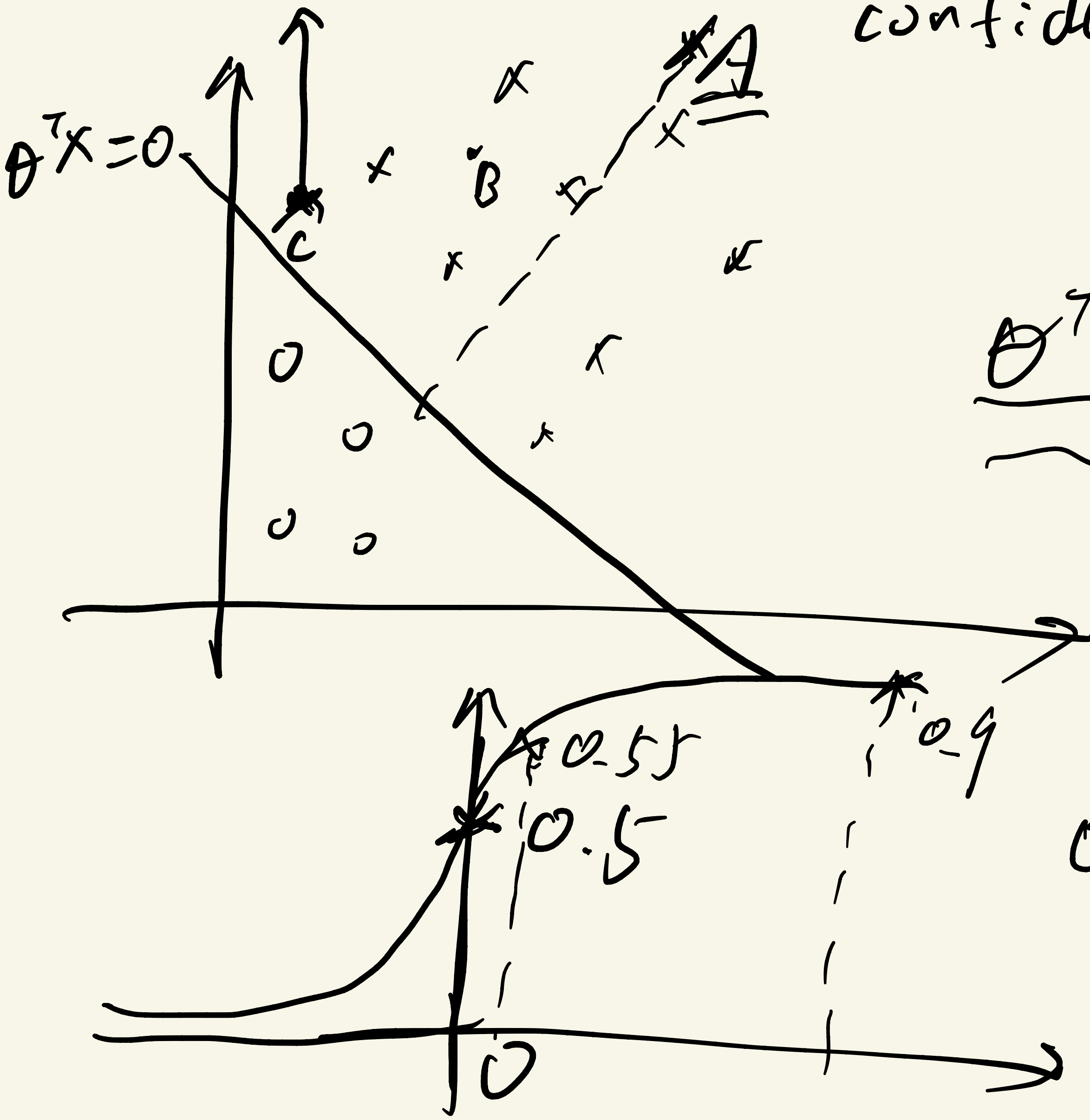
$z^T A z > 0$ positive definite

$z^T A z \leq 0$ negative semi definite

$\langle x, z \rangle \rightarrow K = (x^T z)^K$

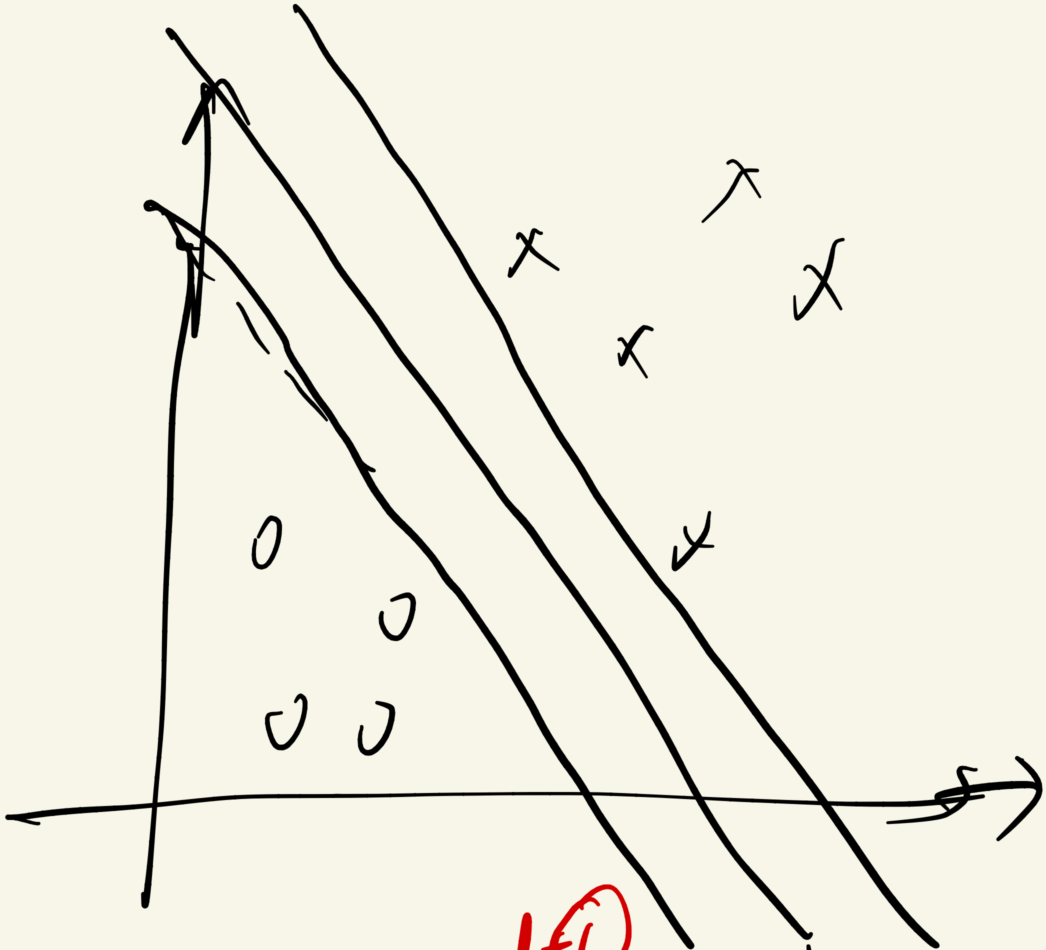
less confident

confident A



$\theta^T x \geq 0$

$\theta^T x >> 0$



$\theta^T x \quad \theta \in R^{d+1}$

$= \theta_2 x_2 + \theta_1 x_1 + \theta_0$

$x = \begin{bmatrix} 1 \\ x_1 \\ x_2 \end{bmatrix}$

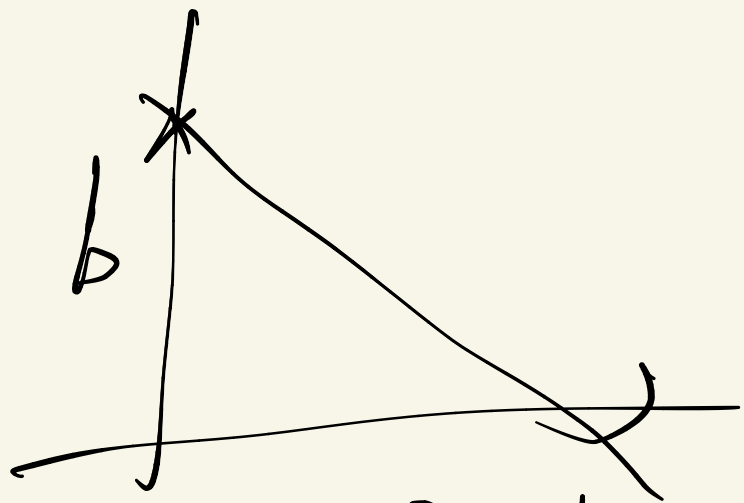
best

$w^T x + b$

θ_0

$w^T x + b$

$y = \underset{\sim}{w} x + \underset{\sim}{b}$



$$\underset{\approx}{g}(w^T x + b) = \begin{cases} 1 & w^T x + b \geq 0 \\ -1 & w^T x + b < 0 \end{cases}$$

$$g(z) = \begin{cases} \text{bool}\left(\dfrac{1}{1+e^{-z}} \geq 0.5\right) \\ \cancel{\text{bool}(} \end{cases}$$

$$\hat{r}^{(i)} = \underbrace{y^{(i)}}_{1}(\underbrace{w^T x^{(i)} + b}_{\geq 0})$$

ith example

Assumption:
training data
is linearly
seperatable

positive: $\hat{r}^{(i)} \geq 0$

negative: $w^T x^{(i)} + b < 0$

$y^{(i)} = -1$

$\hat{r}^{(i)} > 0$

$w \to 2w$

$b \to 2b$

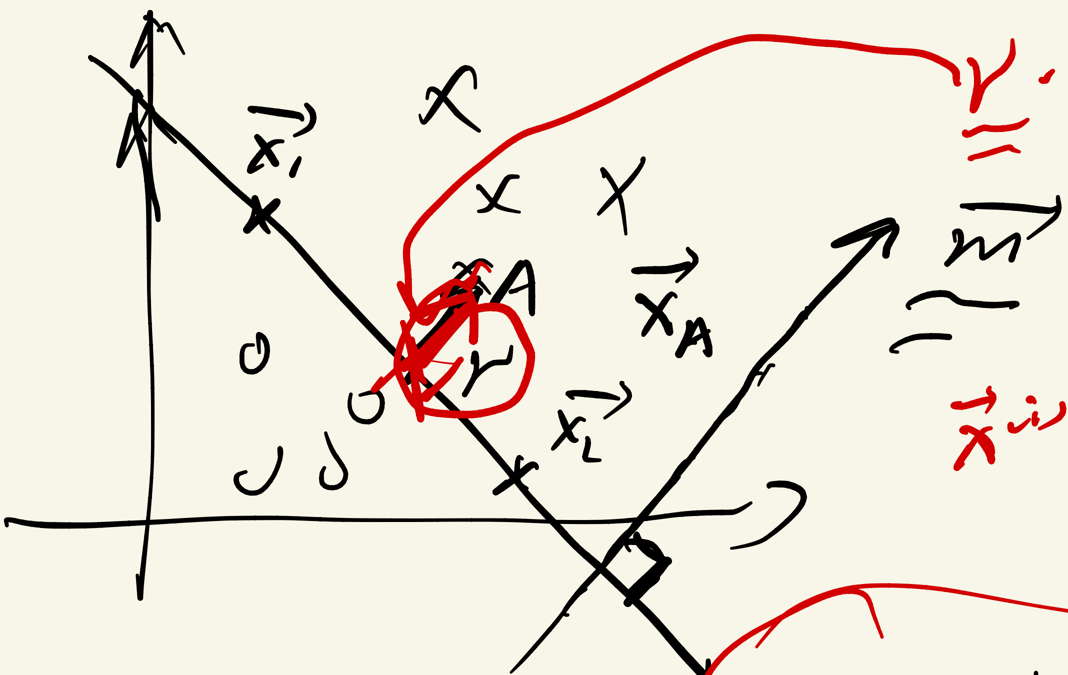
$\|w\| = \sqrt{w^T w}$

$$\left.\begin{array}{l} w^T x + b = 0 \\ 2w^T x + 2b = 0 \end{array}\right\} \text{same}$$

$\hat{r}^{(i)} \to 2\hat{r}^{(i)}$

$$\arg\max_w \frac{1}{\|w\|} = \underbrace{\arg\min_w}\ \|w\|$$

$$= \arg\min_w \|w\|^2$$



$$\gamma \cdot \frac{\vec{w}}{\|\vec{w}\|} = \vec{m}$$

$$\vec{x}^{(i)} - \gamma \cdot \frac{\vec{w}}{\|\vec{w}\|}$$

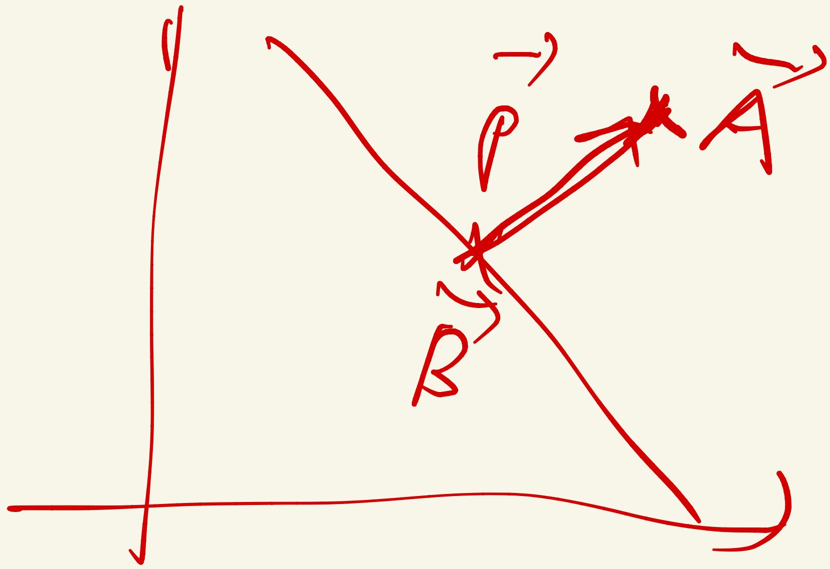
$$w^T x + b = 0 \qquad \vec{w}$$

$\forall\ \vec{x}_1, \vec{x}_2$ that s.t. $w^T x + b = 0$

$$\vec{m} \cdot (\vec{x}_1 - \vec{x}_2) = 0$$

$$\left.\begin{aligned} w^T \vec{x}_1 + b &= 0 \\ w^T \vec{x}_2 + b &= 0 \end{aligned}\right\}$$

$$w^T (\vec{x}_1 - \vec{x}_2) = 0 \qquad \vec{w}$$



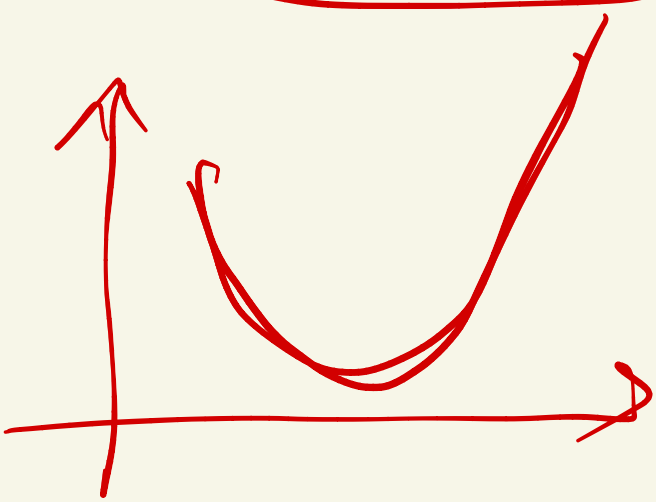
$$\vec{A} - \vec{P} = \vec{B}$$

$$\vec{B} = \left(\underset{\sim}{x}^{(i)} - \gamma^{(i)} \frac{\vec{w}}{\|w\|} \right)$$

$$w^T \vec{B} + b = 0$$

$$\hat{\gamma}^{(i)} = y^{(i)} \left(w^T x^{(i)} + b \right)$$

$$\gamma^{(i)} = \frac{\hat{\gamma}^{(i)}}{\|w\|}$$



$\hat{\gamma}$