TOPICS GO OVER.

① ~~SVM Derivation : Dual Reasoning.~~ ③ ~~Convex Functions.~~

② ~~Design Matrix.~~ ④ ~~PD kernel + Dual.~~

DESIGN MATRIX:

$X \in \mathbb{R}^{n \times d}$ (training examples as rows and features as columns)

$$X = \underbrace{\begin{bmatrix} — & x^{(1)} & — \\ — & x^{(2)} & — \\ & \vdots & \\ — & x^{(n)} & — \end{bmatrix}}_{d \text{ features.}}$$ where $x^{(i)} \in \mathbb{R}^d$.

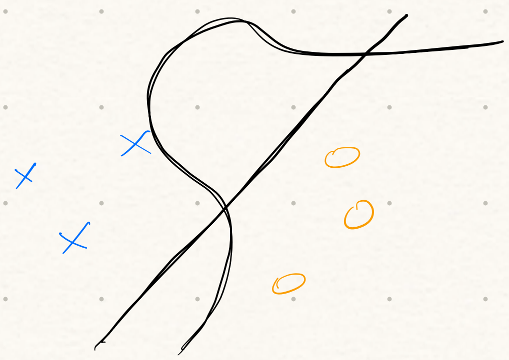
Ex.: Polynomial Regression:

$d=1$: $(\underset{=x^{(1)}}{2}, \underset{=y^{(1)}}{5})$, $(\underset{=x^{(2)}}{-1}, \underset{=y^{(2)}}{0})$, $(\underset{=x^{(3)}}{0}, \underset{=y^{(3)}}{3})$

$X \in \mathbb{R}^{3\times 1}$ $\begin{bmatrix} 2 \\ -1 \\ 0 \end{bmatrix}$ $X' \in \mathbb{R}^{3 \times 3}$ $\begin{bmatrix} 1 & x^{(1)} & (x^{(1)})^2 \\ 1 & x^{(2)} & (x^{(2)})^2 \\ 1 & x^{(3)} & (x^{(3)})^2 \end{bmatrix}$

$$= \begin{bmatrix} 1 & 2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$$

$$= \begin{bmatrix} w_1 + w_2 x_1^{(1)} + w_3 (x_1^{(1)})^2 \\ \vdots \end{bmatrix}$$



② SVM:

PRIMAL: $$\min_{\substack{w \in \mathbb{R}^d \\ w_0 \in \mathbb{R}}} \frac{C}{n} \sum_{i=1}^n \underbrace{\max(1 - y^{(i)}(w^T x^{(i)} + w_0), 0)}_{\varepsilon_i} + \frac{1}{2}\|w\|_2^2.$$

$K(x,z) = x^T z$

DUAL: $$\max_{\alpha \in \mathbb{R}^n} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y^{(i)} y^{(j)} \boxed{(x^{(i)})^T x^{(j)}}$$

$K(x^{(i)}, x^{(j)})$

$= ((x^{(i)})^T x^{(j)})^2$

$= K(x^{(i)}, x^{(j)})$

s.t. $\sum_{i=1}^n \alpha_i y^{(i)} = 0$, $\alpha_i \in [0, \frac{C}{n}]$ $\forall i \in [n]$.

① Dual has n variables ; Primal has $O(d)$ variables to optimize.

② α_i Represents constraints. For SVM: $w^* = \sum_{i=1}^{n} \alpha_i x^{(i)}$.

$$\max_{\alpha} \left(\min_{w, b, \xi} L(\alpha, w, b, \xi) \right)$$

$$\min_{w, b, \xi} \left(\max_{\alpha} L(\alpha, w, b, \xi) \right)$$

$$\frac{\partial}{\partial \alpha_i} \sum_{i=1}^{n} \alpha_i c_i = c_i$$

$\min f(x)$

s.t. $x - 1 = 0$

$f(x) + \alpha(x-1)$

$x - 1 = 0$

X ↓ Matrix.

$\mathcal{X}$ ↓ Input Space. ($\mathbb{R}^d$).

$\longrightarrow x^{(i)} \in \mathcal{X} = \mathbb{R}^d$

$x_j^{(i)} \in \mathbb{R}$

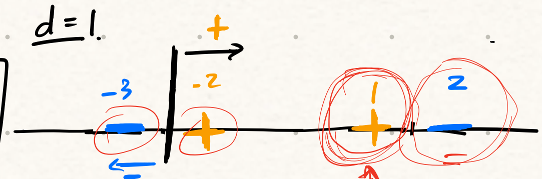
$$\min_{x} \left(\max_{\lambda} L(x, \lambda) \right)$$

$$\boxed{\lambda_i^* g_i(x^*) = 0}$$

SVM (Comp. Slackness): $\boxed{f^*(x) = x^T w^*}$.

$$\boxed{\begin{aligned} &1)\ y^{(i)} f^*(x^{(i)}) > 1 \Rightarrow \alpha_i^* = 0. \\ &2)\ y^{(i)} f^*(x^{(i)}) = 1 \Rightarrow \alpha_i^* = 0 \text{ or } \alpha_i^* > 0 \\ &3)\ y^{(i)} f^*(x^{(i)}) < 1 \Rightarrow \alpha_i^* > 0. \end{aligned}}$$

$d = 1$.



$(-2, +1)$
$(-3, -1)$
$(1, +1)$
$(2, -1)$.

$$\alpha_1^* x^{(1)} + \alpha_2^* x^{(2)} + \underbrace{\cancel{\alpha_3^* x^{(3)}}}_{=0} + \underbrace{\alpha_4^* x^{(4)}}_{>0}$$

$f^*(x) = x^T w^*$

POSITIVE DEFINITE KERNELS.

$K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$.

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{\sigma^2}\right)$$

$K : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$.

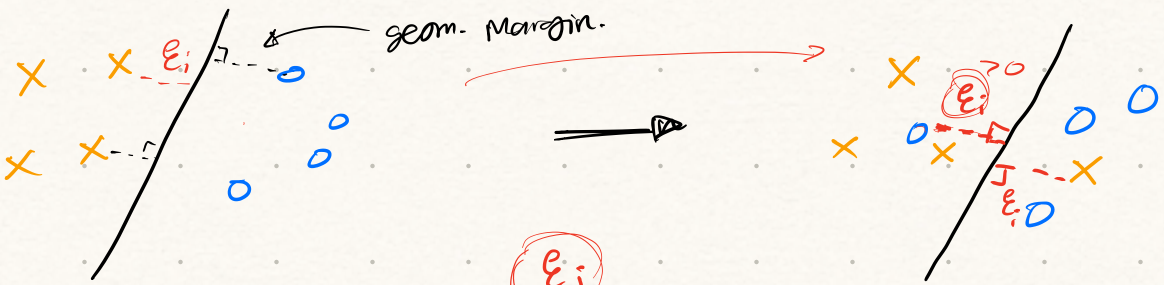
$$\begin{aligned} K(x, z) &= (x^T z)^2. \\ &= (x_1 z_1 + x_2 z_2)^2 \\ &= \langle \phi(x), \phi(z) \rangle \end{aligned}$$

$$\phi(x) = (x_1^2, \sqrt{2} x_1 x_2, x_2^2).$$

$$\boxed{f(x+\delta) \approx f(x) + \nabla f(x)^T\delta} + \frac{\delta^T \nabla^2 f(x)\,\delta}{2} + \dots$$

First order Def.: $\boxed{f(y) \ge \nabla f(x)^T(y-x) + f(x)}$

HARD-MARGIN: only for linearly separable data.



ε_i

SOFT-MARGIN: $\min_{w,b}\; C\sum \underbrace{\max\{0,\, 1-y^{(i)}(w^Tx^{(i)}+w_0)\}}_{\text{hinge}} + \underbrace{\frac{1}{2}\|w\|^2}_{\text{Regularizer}}$

$$1-y^{(i)}(w^Tx^{(i)}+b) \le 0$$

$$1-\underbrace{y^{(i)}(w^T(x^{(i)}+b)}_{\text{Margin}} > 0 \quad \varepsilon_i$$

T/F PROBLEM 5.3:

$\Phi: \mathcal{X} \to \mathcal{X}'$

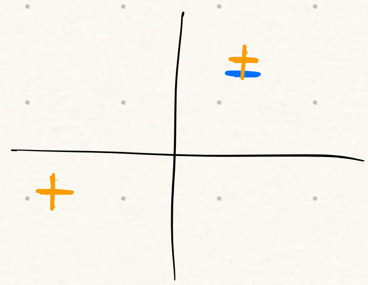
$\Phi(0)$

$(0, +)$

$(0, -)$

$(1, +)$

$P_{Y|X} = Pr(Y=+\mid X=x) = 1/2$



T/F PROBLEM 5.1

$Y = c^*(x) \leftarrow \mathcal{X} = \{0,1\}^5$

$y = \{-1,+1\}$

32

$X = \overset{= c_5}{(0,1,0,1,0)}$ 32 possibilities.

$p_1, \dots, p_{32}$.

$$\Phi(x) = (\mathbb{1}\{x=p_1\}, \mathbb{1}\{x=p_2\}, \dots, \mathbb{1}\{x=p_{32}\})$$

$$w = (1, -1, \dots, 1)$$

$$w^T\Phi(x) =$$

$\mathcal{X} = \{0,1\}^2 \qquad c^*(x) = \begin{array}{l} (00) \to -1 \\ (01) \to +1 \\ (10) \to -1 \\ (11) \to -1 \end{array}$

$$x \xrightarrow{\phi} \phi(x) = (\mathbb{1}\{x=00\},\ \mathbb{1}\{x=01\},\ \mathbb{1}\{x=10\},\ \mathbb{1}\{x=11\})$$

$$w = (-1,\ +1,\ -1,\ -1)$$

$$w^T \phi(x)$$