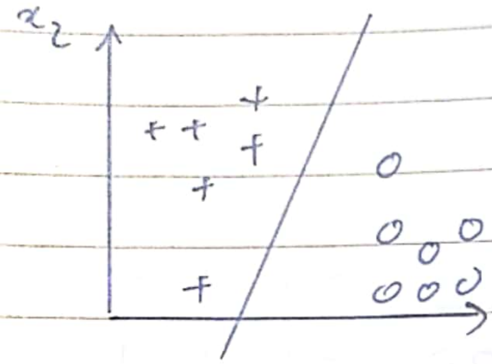
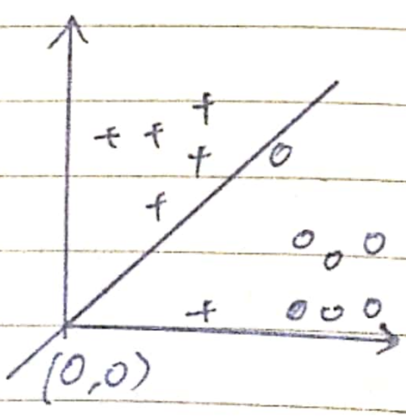
Assignment 3

1. Regularizing Seperate Terms in 2D logistic regression

a) The data are linearly separable so we can find a line that fits the data perfectly.
$\Rightarrow$ 0 classification errors



b) $w_0 = 0 \Rightarrow p(y=1 | x, w) = \sigma(w_1 . x_1 + w_2 . x_2)$
If we take point $(0,0)$ we will see that
$\sigma(w_1 . 0 + w_2 . 0) = \sigma(0) = 0,5$
~~which is the best decision line can have~~ boundary
So the line that passes through $(0,0)$ will be our best decision boundary because any points on one side will have ~~sig~~ sigmoid value $< 0,5$ (class 1) and on the other side will have sigmoid value $> 0,5$ (class 2)
$\Rightarrow$ 1 classification errors in this case

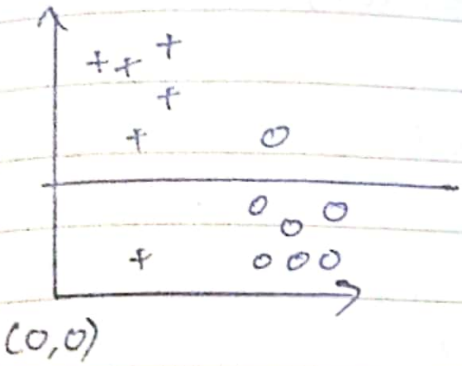


Được quét bằng CamScanner

c) So we will have $w_1 = 0 \Rightarrow p(y=1|x,w) = \sigma(w_0 + x_2 w_2)$

$\Rightarrow$ Decision boundary only depend on x_2

$\Rightarrow$ it is a horizontal line

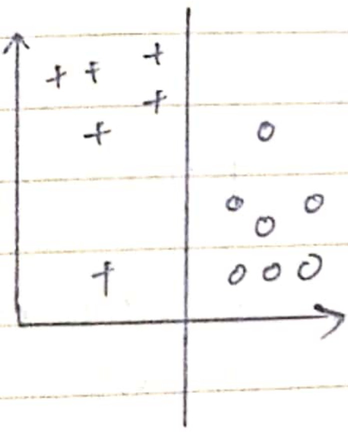


$\Rightarrow$ 2 classification errors

d) we have $w_2 = 0 \Rightarrow p(y=1|x,w) = \sigma(w_0 + x_1 \cdot w_1)$

$\Rightarrow$ Decision boundary only depend on x_1

$\Rightarrow$ it is vertical line



$\Rightarrow$ 0 classification errors

4. Exponential Kernel:

$K(x,z) = \exp(K_1(x,z))$

Using Talor expansion with $a = 0$

$$\Rightarrow \exp(K_1(x,z)) = \exp(0) + \exp'(0).(K_1(x,z) - 0) + \frac{\exp''(0)}{2!}(K_1 - 0) + \ldots$$

$$= 1 + K_1(x,z) + \frac{1}{2}K_1^2(x,z) + \frac{1}{6}K_1^3(x,z) + \ldots$$

Được quét bằng CamScanner

Multiplications and additions of a kernel is also a kernel

$\Rightarrow$ K is a ~~kernel~~ kernel

3. Kernel and corresponding Features:

a) $\phi(x) = \sqrt{c} \cdot \phi_1(x)$

$$\Rightarrow K(x,z) = \phi(x)^T \cdot \phi(z)$$
$$= \sqrt{c} \cdot \phi_1(x)^T \cdot \sqrt{c} \cdot \phi_1(z)$$
$$= c \cdot \phi_1(x)^T \cdot \phi_1(z)$$
$$= c \cdot K_1(x,z)$$

b) $\phi(x) = \begin{bmatrix} \phi_1(x) \\ \phi_3(x) \end{bmatrix}$

$$\Rightarrow K(x,z) = \phi(x)^T \cdot \phi(z)$$
$$= [\phi_1(x)^T \ \phi_3(x)^T] \cdot \begin{bmatrix} \phi_1(z) \\ \phi_3(z) \end{bmatrix}$$
$$= \phi_1(x)^T \cdot \phi_1(z) + \phi_3(x)^T \cdot \phi_3(z)$$
$$= K_1(x,z) + K_3(x,z)$$

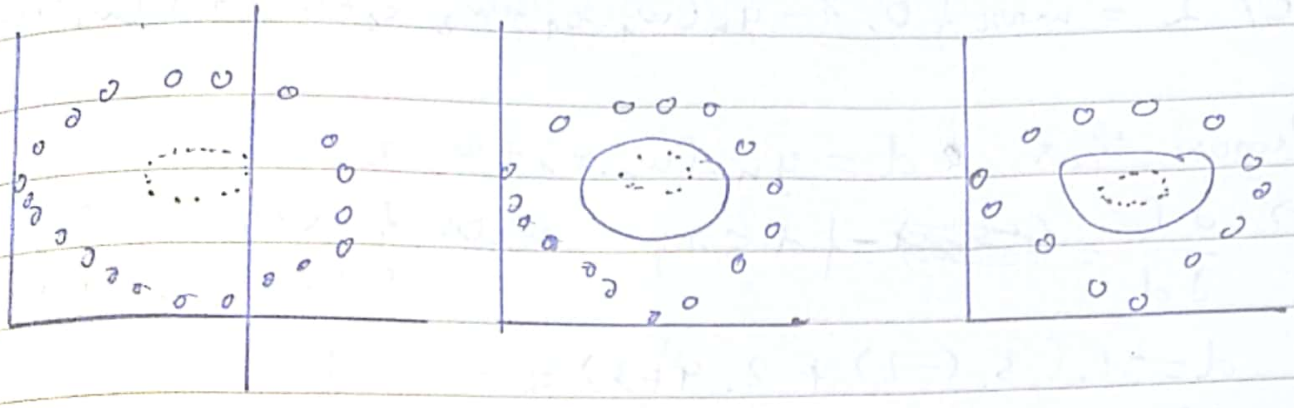
c)

$$\phi(x) = \phi_1(x) \cdot \phi_2(x)$$

$$\Rightarrow K(x,z) = \phi(x)^T \cdot \phi(z)$$
$$= \phi_2(x)^T \cdot \phi_1(x)^T \cdot \phi_1(z) \cdot \phi_2(z)$$
$$= \phi_2(x)^T \cdot \phi_2(z) \cdot \phi_1(x)^T \cdot \phi_1(z)$$
$$= K_2(x,z) \cdot K_1(z,z)$$

Được quét bằng CamScanner

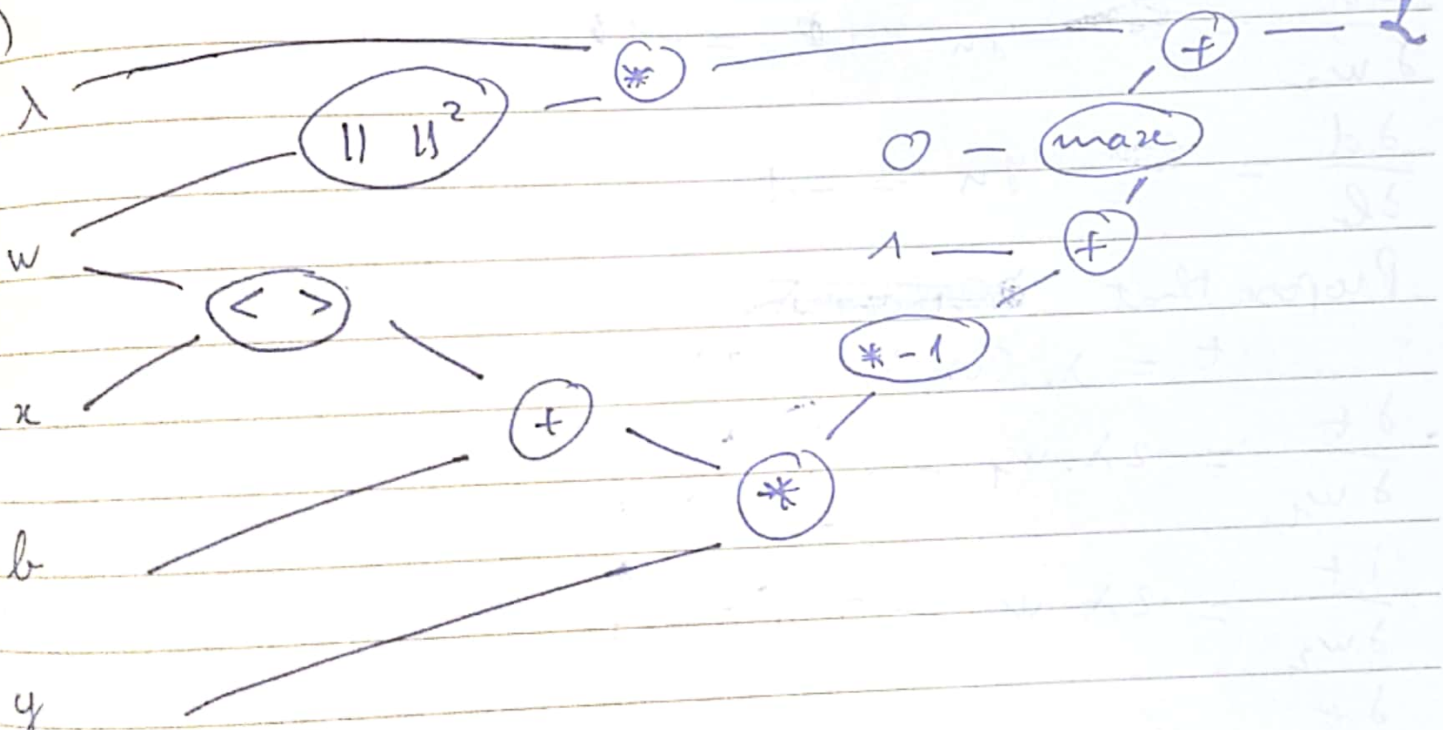
5. SVM with Kernels



linear classifier | polinomial classifier | RBF classifier

6. SVM Back-propagation

a)



b) $\mathcal{L} = \max\{0, 1 - y(\langle \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \rangle + b)\} + \lambda(w_1^2 + w_2^2)$

$= \max\{0, 1 - y.(w_1.x_1 + w_2.x_2 + b)\} + \lambda.(w_1^2 + w_2^2)$

$= \max\{0, 1 + 1.(3.(-2) + 2.4 + 3)\} + 1.(3^2 + 2^2)$

$= \max\{0, 6\} + 13 = 6 + 13 = 19$

Được quét bằng CamScanner

c) $\mathcal{L} = \max\{0, 1 - y_n(w_1.x_1 + w_2.x_2 + b)\} + \lambda.(w_1^2 + w_2^2)$

Propose that $d = y_n.(w_1.x_1 + w_2.x_2 + b)$

$\Rightarrow \dfrac{\partial \mathcal{L}}{\partial d} = -\{d < 1\}$ with $\{d<1\} = 0$ if false, $\{d>1\} = 1$ if true

$d = -1.(3.(-2) + 2.4 + 3) = -5 < 1$

$\Rightarrow \dfrac{\partial \mathcal{L}}{\partial d} = -1$

$\dfrac{\partial d}{\partial w_1} = y_n.x_1 = -1.(-2) = 2$

$\dfrac{\partial d}{\partial w_2} = y_n.x_2 = -1.4 = -4$

$\dfrac{\partial d}{\partial b} = y_n = -1$

Propose that

$t = \lambda.(w_1^2 + w_2^2)$

$\dfrac{\partial t}{\partial w_1} = 2\lambda.w_1 = 2.1.3 = 6$

$\dfrac{\partial t}{\partial w_2} = 2\lambda.w_2 = 2.1.2 = 4$

$\dfrac{\partial t}{\partial b} = 0$

We have:

$\dfrac{\partial \mathcal{L}}{\partial w_1} = \dfrac{\partial \mathcal{L}}{\partial d}.\dfrac{\partial d}{\partial w_1} + \dfrac{\partial t}{\partial w_1} = -1.2 + 6 = 4$

$\dfrac{\partial \mathcal{L}}{\partial w_2} = \dfrac{\partial \mathcal{L}}{\partial d}.\dfrac{\partial d}{\partial w_2} + \dfrac{\partial t}{\partial w_2} = -1.(-4) + 4 = 8$

$\dfrac{\partial \mathcal{L}}{\partial b} = \dfrac{\partial \mathcal{L}}{\partial d}.\dfrac{\partial d}{\partial b} + \dfrac{\partial t}{\partial b} = -1(-1) + 0 = 1$

Được quét bằng CamScanner

d) Learning rate $\alpha = 1$

Step 1:

$$w_{1new} = w_{1old} - \alpha \cdot \frac{\partial \mathcal{L}}{\partial w_1} = 3 - 1.4 = -1$$

$$w_{2new} = w_{2old} - \alpha \cdot \frac{\partial \mathcal{L}}{\partial w_2} = 2 - 1.8 = -6$$

$$b_{new} = b_{old} - \alpha \cdot \frac{\partial \mathcal{L}}{\partial b} = 3 - 1.1 = 2$$

2. SVM

a)

$$\phi(x_1) = [1, \sqrt{2}.0, 0^2]^T = [1, 0, 0]^T$$

$$\phi(x_2) = [1, \sqrt{2}.\sqrt{2}, \sqrt{2}^2]^T = [1, 2, 2]^T$$

~~$\phi(x_1)$ and $\phi(x_2)$~~ $\phi(x_1) \to \phi(x_2) \perp$ hyperplane

$$\phi(x_2) - \phi(x_1) = [0, 2, 2]^T$$

$\Rightarrow w \parallel [0, 2, 2]^T$

~~b) Margin is the distance between $\phi(x_1)$ and $\phi(x_2)$~~

~~$\Rightarrow$ Margin $= \sqrt{0^2 + 2^2 + 2^2}$~~

~~$\Rightarrow$ Margin $= \sqrt{(1-1)^2 + (2-0)^2 + (2-0)^2} = \sqrt{8} = 2\sqrt{2}$~~

b) 2 times margin is the distance between $\phi(x_1)$ and $\phi(x_2)$

$$\Rightarrow \text{Margin} = \frac{\sqrt{(1-1)^2 + (2-0)^2 + (2-0)^2}}{2} = \frac{\sqrt{8}}{2} = \frac{2\sqrt{2}}{2} = \sqrt{2}$$

c) We have

$$\mathcal{L}(w, b, \alpha) = \frac{1}{2}\|w\|^2 + \sum_{i=1}^{N} \alpha_i . (1 - y_i(w^T . \phi(x_i) + b))$$

$$\frac{\partial \mathcal{L}}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{N} \alpha_i \cdot y_i \cdot \phi(x_i)$$

$\forall \alpha_i \geq 0$

Được quét bằng CamScanner

ibis

$$\Rightarrow w = \alpha_1 \cdot y_1 \cdot \phi(x_1) + \alpha_2 \cdot y_2 \cdot \phi(x_2)$$

$$= \alpha_1 . (-1) . \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \alpha_2 . 1 . \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}$$

$$= \begin{bmatrix} -\alpha_1 + \alpha_2 \\ 2\alpha_2 \\ 2\alpha_2 \end{bmatrix}$$

$$\frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^{N} \alpha_i \cdot y_i = 0$$

$$\Leftrightarrow \alpha_1 \cdot y_1 + \alpha_2 \cdot y_2 = 0$$

$$\Leftrightarrow -\alpha_1 + \alpha_2 = 0$$

$$\Leftrightarrow \alpha_1 = \alpha_2$$

$$\Rightarrow w = \begin{bmatrix} 0 \\ 2\alpha_2 \\ 2\alpha_2 \end{bmatrix}$$

we have $\dfrac{1}{\|w\|} = \sqrt{2}$

$$\Leftrightarrow \|w\| = \frac{1}{\sqrt{2}}$$

$$\Leftrightarrow \|w\|^2 = \frac{1}{2}$$

$$\Leftrightarrow (2\alpha_2)^2 + (2\alpha_2)^2 + 0 = \frac{1}{2}$$

$$\Leftrightarrow 8\alpha_2^2 = \frac{1}{2} \Leftrightarrow \alpha_2^2 = \frac{1}{16} \Rightarrow \alpha_2 = \pm\frac{1}{4} \text{ but } \alpha_2 \geq 0$$

$$\Leftrightarrow \alpha_2 = \frac{1}{4} \Rightarrow w = \begin{bmatrix} 0 \\ 1/2 \\ 1/2 \end{bmatrix}$$

~~$w = [0, \frac{1}{2}, \frac{1}{2}]^T$; $w = [0, -\frac{1}{2}, -\frac{1}{2}]^T$ is wrong because we cannot find b in question~~

Được quét bằng CamScanner

d) 2 point closet satisfy

$$y_1 (w^T \cdot \phi(x_1) + b) = 1$$

$$y_2 (w^T \cdot \phi(x_2) + b) = 1$$

$$w^T = [0, \tfrac{1}{2}, \tfrac{1}{2}]$$

$$\Rightarrow \begin{cases} -1 \cdot (0 + b) = 1 \Rightarrow b = -1 \\ 1 \cdot (1 + 1 + b) = 1 \Rightarrow b = -1 \end{cases}$$

So $b = -1$

$$\cancel{w^T = [0, -\tfrac{1}{2}, -\tfrac{1}{2}]}$$

$$\cancel{\Rightarrow \begin{cases} -1 \cdot (0 + b) = 1 \Rightarrow b = -1 \quad (\text{wrong}) \\ 1 \cdot (-1 - 1 + b) = 1 \Rightarrow b = 3 \end{cases}}$$

~~So $b = -1$~~

Được quét bằng CamScanner