SVM-PART-(C)

17 October 2020 17:18

By : VINOD P.

x_1	x_2	$x_1 * x_2$
0	3	0
1	2	2
2	1	2
3	0	0

$\longrightarrow$

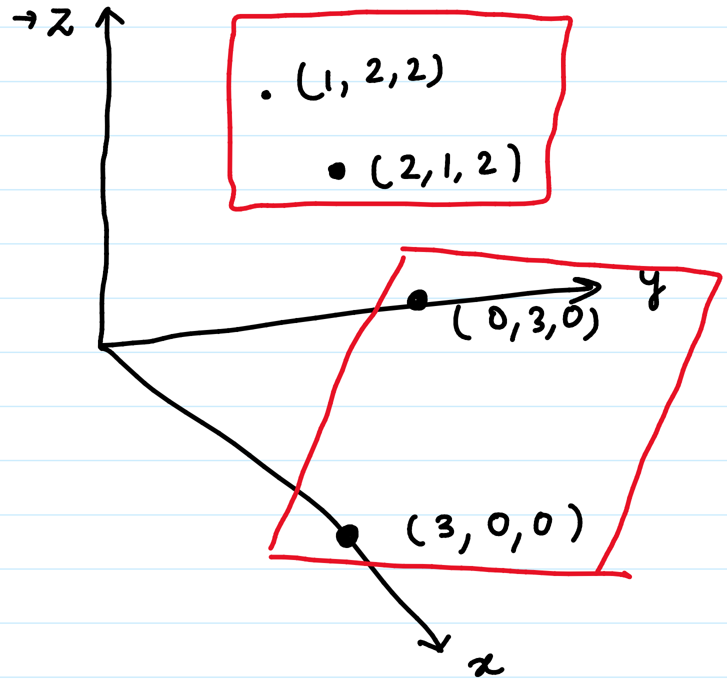
$x_1, x_2, x_1 * x_2$

✓ 0, 3, 0

✓ 1, 2, 2

✓ 2, 1, 2

✓ 3, 0, 0



Kernels

↳ Kernel Tricks

x of dimension d, we can transform x into space using ϕ

$\rightarrow \phi : x \longrightarrow \phi(x)$ K dimensions

$K >> d$

Objective : $\frac{1}{2}\|w\|^2 + C\sum_i \xi_i$

Constraint : $\xi > 0$

$\rightarrow y_i(w \cdot \phi(x) + b) \geq 1 - \xi_i$

$$L = \frac{1}{2}\|w\|^2 + C\sum_i \xi_i - \sum_i \alpha_i\left[y_i(w \cdot \phi(x_i) + b) - 1 + \xi_i\right] - \sum_i \beta_i \xi_i \longrightarrow (A)$$

$\frac{\partial L}{\partial w} = ?$ $\frac{\partial L}{\partial b} = ?$ $\frac{\partial L}{\partial \xi_i}$

$\frac{\partial L}{\partial w} =$ $\boxed{w = \sum_{i=1}^{m} \alpha_i y_i \cdot \phi(x_i)}$

$\frac{\partial L}{\partial b}$: $\boxed{\sum_i \alpha_i y_i = 0}$

$\frac{\partial L}{\partial \xi_i}$: $\boxed{C = \sum \alpha_i + \sum \beta_i}$

$$L = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j \, \phi(x_i) \cdot \phi(x_j)$$

replaced

$K(x_i, x_j)$

$\therefore K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$

Training example x_i : 'd' dimension

$\phi(x_i)$: D where D >> d

$$\sum_{i,j} \alpha_i \alpha_j y_i y_j \, \phi(x_i) \cdot \phi(x_j)$$

d d $\Rightarrow O(d^2)$

$\because$ m observation $O(m^2 . d^2)$

Eg:

	f_1	f_2
	x_1	x_1
	x_1	x_2
	x_2	x_1
	x_2	x_2

x_1 x_2

$\sum_{i,j} \alpha_i \alpha_j y_i y_j \, x_i \cdot x_j$

d d

$x_2 \quad x_1$

$x_2 \quad x_2$

$\phi : x \longrightarrow$ New space (larger Dim)

assume D $(D \gg d)$

$\phi(x_i) \cdot \phi(x_j)$

$\hookrightarrow O(D^2)$

$\hookrightarrow$ for m obs.. $O(m^2 \cdot D^2)$

$$\boxed{K(x_a, x_b) = \phi(x_a) \cdot \phi(x_b)}$$

If we have unknown vector u

$$g(u) = \omega \cdot u + b = \sum \alpha_i y_i \, \phi(x_i) \cdot \phi(u) + b$$

$K(x_i, x_u)$

Kernels

(1) Linear : $K(x_i, x_j) = x_i \cdot x_j$

(2) Polynomial Kernel ; power p

$$K(x_i, x_j) = (1 + x_i \cdot x_j)^p$$

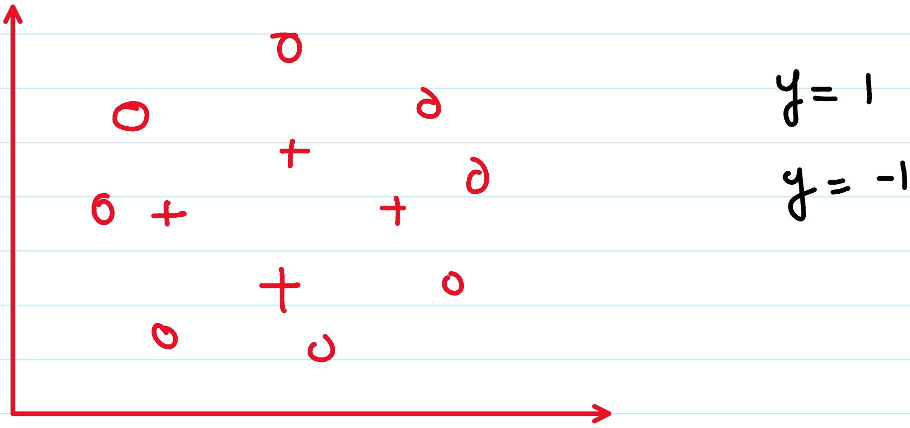
(3) Gaussian (Radial - bais)

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

(4) Sigmoid

$$K(x_i\, x_j) = \tanh(\beta_0 \cdot x_i \cdot x_j + \beta_1)$$

MERCER Cond^n can be called Kernel



$$\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 \cdot x_1^2 + \theta_5 x_2^2 + \cdots$$

$$+ \cdots \geq 0 \rightarrow \text{(B)}$$

observation $\rightarrow$ class 1 ; $y = 1$

θ_j = unknown

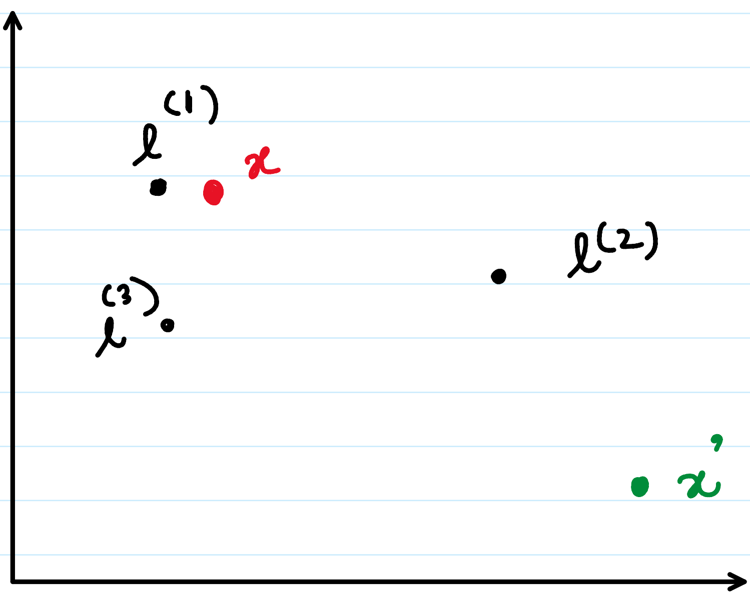
$$h_\theta(x) = \begin{cases} 1 & \text{if } \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Another way for expressing (B) is

$$\theta_0 + \theta_1 f_1 + \theta_2 f_2 + \theta_3 f_3 + \cdots$$

f_i are new features

$f_1 = x_1$, $f_2 = x_2$, $f_3 = x_1 x_2$, $f_4 = x_1^2$ . . .



Idea: Given x $\rightarrow$ find proximity x with

Idea: Given x → find proximity x with $l^{(i)}$

Define f_i

$$f_1 = sim(x, l^{(1)}) = \exp\left(-\frac{\|x - l^{(1)}\|^2}{2\sigma^2}\right)$$

$$f_2 = sim(x, l^{(2)}) = \exp\left(-\frac{\|x - l^{(2)}\|^2}{2\sigma^2}\right)$$

$$f_3 = sim(x, l^{(3)}) = \exp\left(-\frac{\|x - l^{(3)}\|^2}{2\sigma^2}\right)$$

If $x \approx l^{(1)}$

$$f_1 \approx \exp\left(-\frac{0^2}{2\sigma^2}\right) \approx 1$$

$$f_2 \approx \exp\left(-\frac{\text{large value}^2}{2\sigma^2}\right) \approx 0$$

$$f_3 \approx 0$$

Given x: ✓

→ $\theta_0 + \theta_1 f_1 + \theta_2 f_2 + \theta_3 f_3 \geq 0$

$\theta_0 = -0.5$ $\quad\quad$ $\theta_3 = 0$

$\theta_1 = 1$

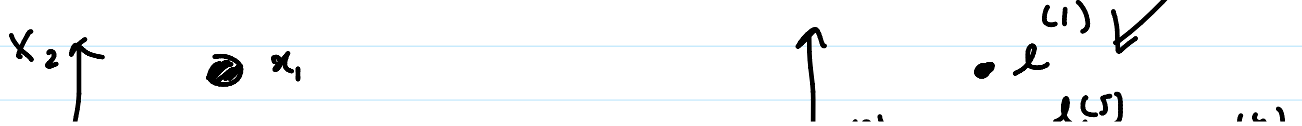
$\theta_2 = 1$

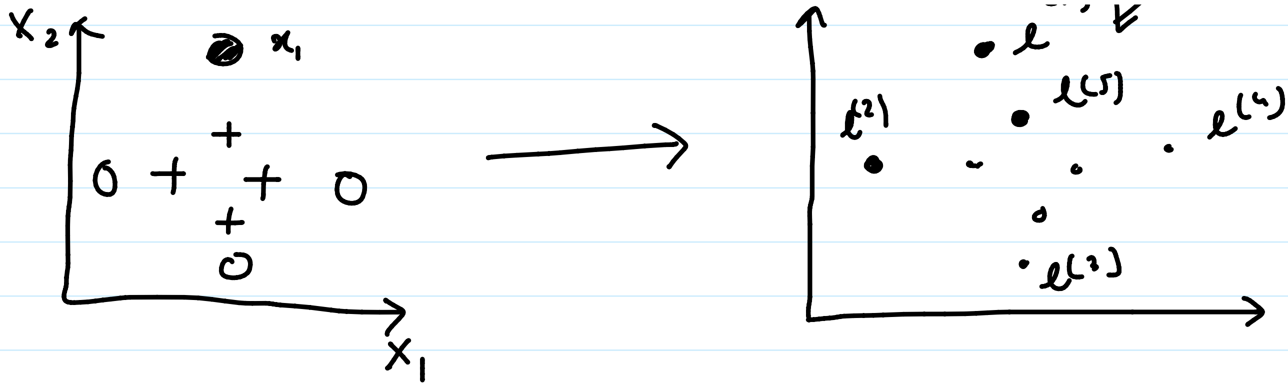
for x: $f_1 = 1$, $f_2 = 0$, $f_3 = 0$

$-0.5 + 1 \times 1 + 1 \times 0 + 0 \times 0$

$-0.5 + 1 = 0.5$ we say x is +ve

How to choose landmark ?





Landmark will put to same location as training example

Given eg x:

$$f_1 = sim(x, l^{(1)})$$

$$f_2 = sim(x, l^{(2)})$$

...

$(x^{(i)}, y^{(i)})$: Train. eg.

$$x^{(i)} \rightarrow \boxed{\begin{matrix} f_1^{(i)} \\ f_2^{(i)} \\ \vdots \\ f_n^{(i)} \end{matrix}} \begin{matrix} = sim(x^{(i)}, l^{(1)}) \\ = sim(x^{(i)}, l^{(2)}) \\ \vdots \\ = sim(x^{(i)}, l^{(n)}) \end{matrix}$$

$$f_i^{(i)} = \ldots sim(x^{(i)}, l^{(i)}) = exp\left(-\frac{0^2}{2\sigma^2}\right) \approx 1$$

$f =$

hypothesis : $x" \rightarrow f \in \mathbb{R}^n$

predict "$y = 1$" if $\theta^T \cdot f \geq 0$