Applied Machine Learning

HW 2

Irfan Ali - CS16BTECH11019

Q1: Reference taken from: www.cs.cmu.edu/~ggordon/svms/new-svms-and-kernels.pdf

Mercer's Condition: $\int dx dy\, f(x)\, K(x,y)\, f(y) \geq 0.$

or in short: $K(x,y) \geq 0$

(a) $K(x,z) = K_1(x,z) + K_2(x,z)$

$\Rightarrow \int dx dy\, f(x) \cdot f(y)\, K(x,z) = \int dx dy\, f(x) \cdot f(y) (K_1(x,z) + K_2(x,z))$

$= \underbrace{\int dx dy\, f(x) K_1(x,z) f(y)}_{\geq 0} + \underbrace{\int dx dy\, f(x) K_2(x,z) f(y)}_{\geq 0}$

$\rightarrow \int dx dy\, f(x) \cdot f(y)\, K(x,z) \geq 0$

$\therefore$ $K(x,z)$ is valid kernel.

(b) $K(x,z) = K_1(x,z)\, K_2(x,z)$

$K_1(x,z) \geq 0$ and $K_2(x,z) \geq 0$

$\Rightarrow K(x,z) \geq 0$

$\Rightarrow \int dx dy\, f(x) f(y) K_1(x,z) \geq 0$ and $\int dx dy\, f(x)\, K_2(x,z) f(y) \geq 0$

$\Rightarrow \int dx dy\, f(x) \cdot f(y)\, K(x,z) \geq 0$ $\therefore$ $K(x,z)$ is valid Kernel

(c) $K(x,z) = h(K_1(x,z))$ h is polynomial with positive co-effs.

let $K(x,z) = \sum_{i=0}^{q} a_i K_1(x,z)^i$

$\forall a_i > 0$ and also $K_1(x,z) \geq 0$.

$\Rightarrow K(x,z) \geq 0$

$$\int dx\,dy\, f(x) K(x,z) f(y) = \int dx\,dy\, f(x) a_0 f(y) + \int dx\,dy\, f(x) a_1 K_1(x,z) f(y)$$
$$+ \int dx\,dy\, f(x) a_2 (K_1(x,z))^2 f(y) \ldots\ldots$$

All the terms on RHS are non-negative.

$\therefore$ LHS is also non negative.

$\therefore K(x,z)$ is valid kernel.

(d) $k(x,z) = \exp(k_1(x,z))$

$K_1(x,z) \geq 0 \Rightarrow \exp(K_1(x,z)) \geq 1 \rightarrow \exp(k_1(x,z)) \geq 0$

$\therefore K(x,z) \geq 0$.

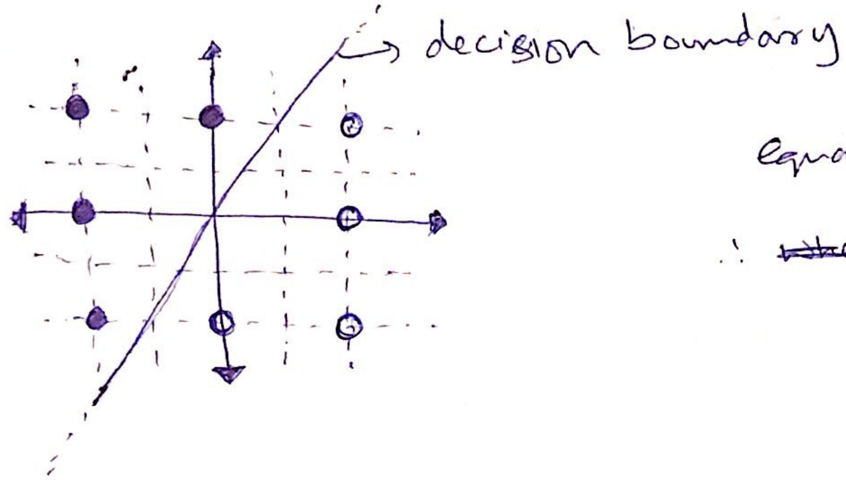
$\therefore K(x,z)$ is a valid Kernel.

(e) $K(x,z) = \exp\left(\frac{-\|x-z\|_2^2}{\sigma^2}\right)$

This Kernel is gaussian RBF Kernel which is already proven to be valid.

$\therefore K(x,z)$ is a valid Kernel.

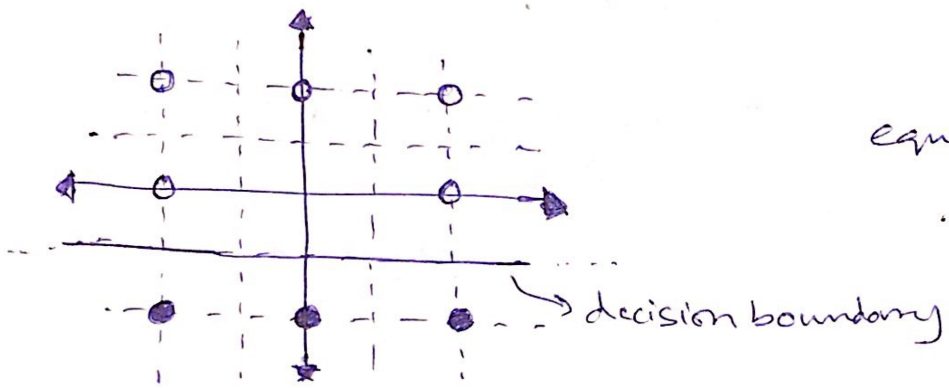
2.



equation ≡ $y = 2x$

$\therefore$ ~~the~~ weight $= 2$

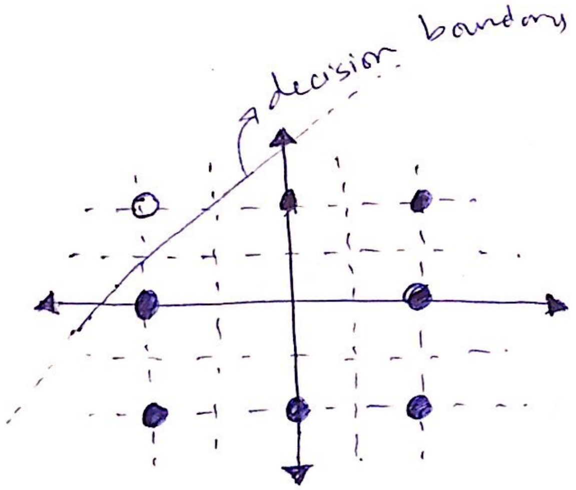
bias $= 0$



equation ≡ $y = -1$

$\therefore$ weight $= 0$

bias $= -1$



equation ≡ $y = x + 3$

$\therefore$ weight $= 1$

bias $= 3$

3 let W_1 and W_2 be the weights.

$$\text{Error} = \frac{1}{2}\sum_{i=1}^{n}\left(x_{1i}W_1 + x_{2i}W_2 - y_i\right)^2$$

where n is the number of samples. $((x_{1i}, x_{2i}), y_i)$

$\Rightarrow$ ~~Eg~~ $X_1 = [1 \;\; 1] \quad y_1 = 1$

$X_2 = [1 \;\; -1] \quad y_2 = -1$

$$\Rightarrow E = \frac{1}{2}(w_1 + w_2 - 1)^2 + \frac{1}{2}(w_1 - w_2 + 1)^2$$

$$\frac{\partial E}{\partial w_1} = 0 \quad \text{and} \quad \frac{\partial E}{\partial w_2} = 0$$

$$\Rightarrow \frac{\partial E}{\partial w_1} = (w_1 + \cancel{w_2} - \cancel{1}) \times 1 + (w_1 - \cancel{w_2} + \cancel{1}) \times 1 = 0$$

$$\Rightarrow 2w_1 = 0 \Rightarrow \boxed{w_1 = 0} \rightarrow (1)$$

$$\frac{\partial E}{\partial w_2} = (w_1 + w_2 - 1) \times 1 + (w_1 - w_2 + 1) \times -1 = 0$$

$$\Rightarrow 2w_2 - 2 = 0 \Rightarrow \boxed{w_2 = 1} \rightarrow (2).$$

For Curvature, calculate $\frac{\partial^2 E}{\partial w_1^2}$, $\frac{\partial^2 E}{\partial w_2^2}$

$$\frac{\partial^2 E}{\partial w_1^2} = \frac{\partial}{\partial w_1}\left(\frac{\partial E}{\partial w_1}\right) = \frac{\partial}{\partial w}(2w_1) = 2 > 0 \rightarrow (3)$$

$$\frac{\partial^2 E}{\partial w_2^2} = \frac{\partial}{\partial w_2}\left(\frac{\partial E}{\partial w_2}\right) = \frac{\partial}{\partial w}(2w_2 - 2) = 2 > 0 \rightarrow (4)$$

as both ~~E~~ (3) and (4) are >0, the curve increases from minima on both the axes. So the curve is upwards.

Also $(0, 1)$ is the minima.

b) $\frac{\partial^2 E}{\partial w_1^2} = 2$, $\frac{\partial^2 E}{\partial w_2^2} = 2$.

$$\frac{\partial^2 E}{\partial w_1 \partial w_2} = \frac{\partial}{\partial w_1}\left(\frac{\partial E}{\partial w_2}\right) = \frac{\partial}{\partial w_1}(2w_2 - 2) = 0$$

hessian eigenvalues = $\begin{bmatrix} \frac{\partial^2 E}{\partial w_1^2} & \frac{\partial^2 E}{\partial w_1 \partial w_2} \\ \frac{\partial^2 E}{\partial w_1 \partial w_2} & \frac{\partial^2 E}{\partial w_2^2} \end{bmatrix}$

hessian eigenvalues. = $\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$