ASSIGNMENT - 3 LUCKY SAHANI

3\. a) If K_1 is a Kernel on X, then $K(x,z) = e^{K_1(x,z)}$ is also a kernel.

Now, we know the 4 properties : (Given K_1 is a kernel)

1) $\alpha_1 K_1$ is a kernel funct^n $\forall \alpha_1 > 0$
2) If K_1 & K_2 are kernel functions, the $K_1(x,z) \cdot K_2(x,z)$ is also a kernel function
3) If K_1 & K_2 are kernel functions, then $K_1(x,z) + K_2(x,z)$ is a kernel function too.
4) $K(x,y) = c$ is a kernel function

Now, the expansion of exp series is :-

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$$

$$\therefore e^{K_1(x,z)} = 1 + K_1(x,z) + \frac{K_1(x,z)^2}{2!} + \ldots$$

We know $K(x,y) = 1$ is a kernel function (Prop 4)

Now, $K_1(x,z)$ is a kernel function. $K_1 \times K_1(x,z)$ is also a kernel function

$\therefore (K_1(x,z))^i$ is a kernel f^n (Prop 2)

$\alpha_i \times \underbrace{(K_1(x,z))^i}$ is a kernel function (Prop 1)

$\sum_{i=1}^{\infty} \alpha_i (K_1(x,z))^i$ is also a kernel function (Prop 3)

Scanned by CamScanner

Using the points above, we can say that $e^{\alpha_1 (x,z)}$ is a kernel function

b) $K(x,z) = \underbrace{e^{\|x\|^2 + \|z\|^2}}_{K_1} \cdot \underbrace{x^T z \,/\, \|x\|^2 * \|z\|^2}_{K_2}$

Now, $K_1 = e^{\|x\|^2 + \|z\|^2}$

Let $\phi = e^{\|x\|^2}$, then $\phi(x) = e^{\|x\|^2}$, $\phi(z) = e^{\|z\|^2}$

$<\phi(x) \cdot \phi(z)> = e^{\|x\|^2} \cdot e^{\|z\|^2}$

$\therefore$ K_1 is a kernel function.

Now, $K_2 = \dfrac{x^T z}{\|x\|^2 \|z\|^2} = \left(\dfrac{x}{\|x\|^2}\right) \cdot \left(\dfrac{z}{\|z\|^2}\right)$

$= <B(x) \cdot B(z)>$

K_2 is a kernel function

Since, product of two kernel function is a kernel function (By prop. 2), $k(x,z)$ is a kernel function.

---

4) a) minimize $L = \sum_{i=1}^{n} \varepsilon_i^2$

$y_i - w^T x_i = \varepsilon_i \quad \forall i = 1 \ldots n$

$\|w\|_2^2 \leq B$

$\|w\|^2 - B^2 \leq 0$

Now, the $g(x)$ will be $||W||^2 - ||B||^2$ as $g(x) \leq 0$ is there.

Also, $h_i(x) = 0$ was a condition.

We know that

$$y_i - W^T x_i = \varepsilon_i$$

$$\therefore \underbrace{y_i - W^T x_i - \varepsilon_i}_{\to h_i(x)} = 0 \quad \forall i = 1, 2 ..$$

Optimisation problem

$$L(W, \varepsilon, \mu, \lambda) = \sum_{i=1}^{n} \varepsilon_i^2 + \sum_{i=1}^{n} \mu_i h_i(x) + g(x)\lambda$$

$$= \sum_{i=1}^{n} \varepsilon_i^2 + \sum_{i=1}^{n} \mu_i (y_i - W^T x_i - \varepsilon_i) + \lambda(||W||^2 - B^2)$$

Now, $L_d = \min_{x} L(x, \bar{\alpha}, \bar{\beta})$

Here $\alpha = \mu$, & $\bar{\beta} = \lambda$. $\max_{\bar{\alpha}, \bar{\beta}, \alpha \geq 0} L_d(\bar{\alpha}, \bar{\beta})$ is called the dual problem.

Following was the result of KKT theorem!

$$\frac{\partial}{\partial x} L(x^*, \bar{\alpha}^*, \bar{\beta}^*) = 0$$

$$\frac{\partial}{\partial \beta_j} L(x^*, \bar{\alpha}^*, \bar{\beta}^*) = 0 \quad \text{for all } j$$

Scanned by CamScanner

We will subsite the values of α & β according

$$\frac{\partial L}{\partial w} = 0 \Rightarrow -\sum \mu_i x_i + \lambda(2w) = 0$$

$$\therefore w = \sum_{i=1}^{n} \frac{\mu_i x_i}{2\lambda} \qquad \text{--- } ①$$

$$\frac{\partial L}{\partial x_i \to \varepsilon_i} = 0 \Rightarrow 2\varepsilon_i - \mu_i = 0$$

$$\therefore \varepsilon_i = \frac{\mu_i}{2} \qquad \text{--- } ②$$

So,

$$L_d(\mu, \lambda) = \sum \left(\frac{\mu_i}{2}\right)^2 + \sum \mu_i \left(y_i - \frac{1}{2\lambda}\underbrace{(\sum \mu_i x_i)^T}_{w^T} x_i - \frac{\mu_i}{2}\right)$$

$$+ \lambda\left((\sum \mu_i x_i)^T \cdot (\sum \mu_i x_i) - B^2 \cancel{\times 4\lambda^2}\right) \cdot \frac{1}{(2\lambda)^2}$$

$$= \frac{\sum \mu_i^2}{4} + \sum \mu_i y_i - \frac{1}{2\lambda} \sum \mu_i (\sum \mu_i x_i)^T x_i$$

$$- \sum \mu_i \cdot \frac{\mu_i}{2} + \frac{1}{4\lambda} \sum_j \sum_i (\mu_i x_i)^T (\mu_j x_j)$$

$$- \frac{1}{\cancel{4\lambda}} \times \cancel{4} \lambda^{\cancel{2}} \times B^2$$

$$= -\frac{\sum \mu_i^2}{4} + \sum \mu_i y_i - \frac{1}{4\lambda} \sum_i \sum_j (\mu_i x_i)^T (\mu_j x_j)$$

$$- \lambda B^2$$

Scanned by CamScanner

$$\therefore L_d(\mu,\lambda) = -\frac{\sum \mu_i^2}{4} + \sum \mu_i y_i - \frac{1}{4\lambda}\sum\sum (\mu_i x_i)^T (y_j x_j) - \lambda B^2$$

↳ Dual problem

Now, we need to maximize the above

∴ we need to find $\sum \mu_i$ for which the $\frac{\partial L_d}{\partial U}$ is 0

Let $U = \sum \mu_i$,

$$\frac{\partial L_d}{\partial U} = -\frac{(2U)}{4} + \frac{\partial U}{\partial U}\cdot \bar{y} - \frac{2U x^T \cdot x}{4\lambda} = 0$$

$$\bar{U} = \bar{y} * 2\lambda / (\lambda + x^T x) \quad \text{——} \ (3)$$

Also, $\frac{\partial L_d}{\partial \lambda} = 0$

$$\therefore \frac{1}{4\lambda^2}\left(\sum\sum (\mu_i x_i)^T (\mu_j x_j) - B^2 = 0\right.$$

$$\sum\sum (\mu_i x_i)^T (\mu_j \cdot x_j) = 4\lambda^2 B^2$$

Now $w = \frac{\sum U_i x_i}{2\lambda}$ (From (1))

$$\bar{U}$$

$$w = \frac{\sum \mu_i \, x_i}{2\lambda}$$

Get the value from (3)

$$w = \frac{x^T}{2\lambda} \cdot \bar{y}\left[\frac{1}{2\lambda} + \frac{x^T \cdot x}{2\lambda}\right]$$

Scanned by CamScanner

$$w = \frac{x^T}{2\lambda} \cdot \bar{y} \cdot 2\lambda \Big/ (\lambda + x^T x)$$

$$\boxed{w = x^T \bar{y} \Big/ (\lambda + x^T x)}$$

b) As seen before,

$$w = \sum \mu_i \cdot x_i \,/\, 2\lambda$$

$\therefore$ It is equivalent to the support vectors in SVM's

(c) SVM's is much more geometrically optimized where we can find the best separating hyper-plane & also, SVM is faster for the kernel space.

---

3 (c) $K(x, z) = \sum_{i=1}^{d} \min(|x_i|, |z_i|)$ is a kernel

If we can prove $\min(|x_i|, |z_i|)$ is a kernel then $K(x, z)$ will be a kernel as by prop 3, sum of kernels is also a kernel

Let V_i be the max value i^{th} feature can attain for vector x

Define $F(x) = \begin{cases} 1 & \text{if } h \le x_i \\ 0 & x_i < h \le V_i \end{cases}$

consider,

$$F(x) \cdot F(y) = \sum_{j=1}^{v_i} F(x)_j \, F(y)_j$$

$$F(x) \cdot F(y) = \sum_{j=1}^{x_i} F(x)_j \, F(y)_j + \sum_{j=x_i+1}^{v_i} F(x)_j \, F(y)_j$$

Now, we will see cases for x_i

If x_i is between j & y_i, then both F are 1.

If x_i is less then j, then one F is 0.

$\therefore$ Product is 0.

$$\therefore F(x) \cdot F(y) = \sum_{j=1}^{x_i} 1 \cdot 1 = x_i.$$

$\therefore$ $K_i(x, y) = \langle F(x), F(y) \rangle$ where F take vectors from $(Z^+)^D \rightarrow (Z)^{v_i}$ where v_i is the max value

$\therefore$ $K(x, z)$ is a kernel