Ques 3 (a) $\bar{x}, \bar{z} \in \mathcal{X} = \mathbb{R}^d$

if K_1 is a kernel on $\mathcal{X}$, then

$K(\bar{x}, \bar{z}) = e^{K_1(\bar{x}, \bar{z})}$ is a also a kernel.

Sol^n We know,

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots$$

Define $K_n(\bar{x}, \bar{z}) = 1 + K_1(\bar{x}, \bar{z}) + \frac{[K_1(\bar{x}, \bar{z})]^2}{2!} + \frac{[K_1(\bar{x}, \bar{z})]^3}{3!} + \cdots \frac{[K_1(\bar{x}, \bar{z})]^n}{n!}$

Using property "product of two kernels is a kernel",

$[K_1(\bar{x}, \bar{z})]^i$ is a kernal $\forall i = 1, 2, 3 \cdots$

Using property, "sum of two kernels is a kernel",

$$1 + K_1(\bar{x}, \bar{z}) + \frac{[K_2(\bar{x}, \bar{z})]^2}{2!} + \cdots + \frac{(K_1(\bar{x}, \bar{z}))^n}{n!} \text{ is a kernel.}$$

$\Rightarrow$ $K_n(\bar{x}, \bar{z})$ is a kernel.

$\Rightarrow$ $\mu^T K_n(\bar{x}, \bar{z}) \bar{\mu} \geq 0 \quad \forall \mu \in \mathbb{R}^d$

for a converging sequence, if $a_n \geq 0 \ \forall n$

then $\lim_{n \to \infty} a_n = a$ is also greater than equal to 0

Hence, $\lim_{n \to \infty} a_n = a \geq 0$

$$\Rightarrow \lim_{n \to \infty} \left(\mu^T K_n(\bar{x}, \bar{z}) \bar{\mu} \right) \geq 0$$

$$\Rightarrow \mu^T \left(\lim_{n \to \infty} K_n(\bar{x}, \bar{z}) \right) \bar{\mu} \geq 0$$

$$K(\bar{x}, \bar{z}) = \lim_{n \to \infty} K_n(\bar{x}, \bar{z}) = e^{K_1(\bar{x}, \bar{z})}$$

$\Rightarrow K(\bar{x}, \bar{z})$ is a kernel.

classmate

Date

Page

(b) $K(\bar{x},\bar{z}) = e^{(\|x\|^2 + \|z\|^2)} \cdot \left(\frac{\bar{x}^T \bar{z}}{\|x\|^2\|z\|^2}\right)$

Rewrite as

$$K(\bar{x},\bar{z}) = K_1(\bar{x},\bar{z})\, K_2(\bar{x},\bar{z})$$

$$K_1(\bar{x},\bar{z}) = e^{\|x\|^2} e^{\|z\|^2} \qquad K_2(\bar{x},\bar{z}) = \frac{\bar{x}^T \bar{z}}{\|\bar{x}\|^2 \|\bar{z}\|^2}$$

We will show $K_1(\bar{x},\bar{z})$ and $K_2(\bar{x},\bar{z})$ are kernels, then product of two kernels is a kernel, hence $K(\bar{x},\bar{z})$ is a kernel.

$\underline{K_1(\bar{x},\bar{z})}$: $e^{\|x\|^2} e^{\|z\|^2}$

define $f(\bar{x}) = e^{\|x\|^2}$

$$K_1(\bar{x},\bar{z}) = f(\bar{x}) f(\bar{z})$$

for any $\bar{\mu} \in \mathbb{R}^d$, $\mu^T K_1(\bar{x},\bar{z})\, \bar{\mu}$

$$= \sum_{i=1}^{d} \sum_{j=1}^{d} \mu_i \mu_j \left(f(\bar{x}_i) f(\bar{x}_j) \right)$$

$$= \sum_{i=1}^{d} \mu_i f(\bar{x}_i) \sum_{j=1}^{d} \mu_j f(\bar{x}_j)$$

$$= \sum_{i=1}^{d} \left(\mu_i f(\bar{x}_i) \right)^2 \geq 0$$

hence $K_1(\bar{x},\bar{z})$ is a kernel.

$$K_2(\bar{x}, \bar{z}) : \frac{\bar{x}^T \bar{z}}{\|x\|^2 \|z\|^2}$$

define $\quad \phi(\bar{x}) = \dfrac{\bar{x}}{\|x\|^2}$

$$\phi(\bar{x})^T \phi(\bar{z}) = \left(\frac{\bar{x}}{\|\bar{x}\|^2}\right)^T \left(\frac{\bar{z}}{\|\bar{z}\|^2}\right)$$

$$= \frac{\bar{x}^T \bar{z}}{\|\bar{x}\|^2 \|\bar{z}\|^2} = K_2(\bar{x}, \bar{z})$$

Hence, $K_2(\bar{x}, \bar{z}) = \langle \phi(\bar{x}), \phi(\bar{z}) \rangle$

$\Rightarrow K_2(\bar{x}, \bar{z})$ is a kernel

Now, $K_1(\bar{x}, \bar{z}) \cdot K_2(\bar{x}, \bar{z})$ is also a kernel.

$\Rightarrow K(\bar{x}, \bar{z})$ is a kernel.

(c) $K(\bar{x},\bar{z}) = \sum_{i=1}^{d} \min(|\bar{x}_i|, |z_i|)$

let , $\bar{x}, \bar{z} \in Z^d = \mathcal{X}$

We will try to find a mapping $\phi(\bar{x})$ s.t.

$$K(\bar{x},\bar{z}) = \langle \phi(\bar{x}), \phi(z) \rangle$$

If such a mapping exist, then $K(\bar{x},\bar{z})$ satisfies $\bar{\mu}^T K(\bar{x},\bar{z})\bar{\mu} \geq 0 \quad \forall \bar{\mu} \in \mathcal{X}$ and K is a kernel

Mapping ϕ:

let N be the max value from all components of n vectors ($\in \mathcal{X}$) used to obtain K.

Define ϕ to be $\phi : \mathcal{X} \rightarrow \{0,1\}^{N \times d}$

let $\bar{x} = \{x_1, x_2, \ldots, x_d\}$

then

$$\phi(\bar{x}) = \{ \overbrace{1, 1, \ldots, 1}^{|x_1|}, \overbrace{0, 0 \ldots, 0}^{N-|x_1|}, \underbrace{1, 1, \ldots 1}_{|x_2|}, \underbrace{0, 0 \ldots 0, 0}_{N-|x_2|}, \ldots \underbrace{1, 1, 1, 1, 1}_{|x_d|}, \underbrace{0 \ldots 0}_{N-|x_d|} \}$$

Now $\phi(x)^T \phi(z) = \sum_{i=1}^{d} \min(|x_i|, |z_i|)$

Hence

$$\underline{k(\bar{x}, \bar{z}) = \langle \phi(\bar{x}), \phi(z) \rangle}.$$

and hence $k(\bar{x}, \bar{z})$ is a kernel.

classmate

Date ______

Page ______

Ques 4 $\min_{\overline{w}, \overline{\xi}} \left\{ \sum_{i=1}^{n} \xi_i^2 \right.$

constraints:-

$$① \quad y_j - \overline{w}^T \overline{x_j} - \xi_i = 0 \quad \forall j = 1, \dots, n$$

$$② \quad \|\overline{w}\|_2^2 - B^2 \le 0$$

$y_j \in \mathbb{R}, \ \overline{x_i} \in \mathbb{R}^L$

Sol^n (a) obtaining Langrangian

$$L(\overline{w}, \overline{\xi}, \alpha, \overline{\beta}) = \sum_{i=1}^{n} \xi_i^2 + \alpha\left(\|\overline{w}\|^2 - B^2\right) + \sum_{j=1}^{n} \beta_j \left(y_j - \overline{w}^T \overline{x_j} - \xi_j\right)$$

Applying stationary conditions:

1. $$\frac{\partial L}{\partial \overline{w}} = 2\alpha \overline{w} - \sum_{j=1}^{n} \beta_j \overline{x_j} = 0$$

$$\Rightarrow \boxed{\overline{w} = \frac{1}{2\alpha} \sum_{j=1}^{n} \beta_j \overline{x_j}}$$

let X be data matrix $[X]_{n \times L} = \begin{bmatrix} \overline{x_1}^T \\ \overline{x_2}^T \\ \vdots \\ \overline{x_n}^T \end{bmatrix}$, $\overline{x_i} \in \mathbb{R}^L$

$$\boxed{\overline{w} = \frac{X^T \overline{\beta}}{2\alpha}} \quad \text{—— ①}$$

Vikas Jain

2. $\frac{\partial L}{\partial \bar{\xi}} = 2\bar{\xi} - \bar{\beta} = 0$

$$\Rightarrow \boxed{\bar{\xi} = \frac{\bar{\beta}}{2}} \quad \text{--- (2)}$$

from (1) & (2), modified langrangian is

$$L(\bar{w}, \bar{\xi}, \alpha, \bar{\beta}) = \frac{1}{4}\|\bar{\beta}\|^2 + \alpha\left(\frac{1}{4\alpha^2}\bar{\beta}^T X X^T \bar{\beta} - B^2\right) + \bar{\beta}^T\left(\bar{y} - \frac{X X^T \bar{\beta}}{2\alpha} - \frac{\bar{\beta}}{2}\right)$$

Hence, the dual optimization problem becomes.

$$\boxed{L_D = \min_{\alpha, \bar{\beta}} \left(\frac{1}{4}\|\bar{\beta}\|^2 + \alpha\left(\frac{1}{4\alpha^2}\bar{\beta}^T X X^T \bar{\beta} - B^2\right) + \bar{\beta}^T\left(\bar{y} - \frac{X X^T \bar{\beta}}{2\alpha} - \frac{\bar{\beta}}{2}\right)\right) \quad \alpha, \bar{\beta} \geq 0}$$

Solving above dual problem, will give $\alpha, \bar{\beta}$ and from $\alpha, \bar{\beta}$, we can find out

$\bar{w}^*, \bar{\xi}^*$.

(b) yes, this problem has equivalent of support vectors as in SVM.

The points from the $\mathcal{L}$ which determines the regression line ($y_i = \overline{W}^T x_i$) are the support vectors.

(i) All support vectors $\overline{x}_i$ lie on the regression line and satisfy

$$\boxed{y_i - \overline{W}^{*T} \overline{x}_i = 0 \quad \text{and} \quad \beta_i = 0}$$

hence the ~~support~~ vectors $\overline{x}_i$ for which $\beta_i = 0$ are support vectors as

$$\beta_i = 0 \Rightarrow \xi_i = 0 \Rightarrow y_i - \overline{W}^{*T} \overline{x}_i = 0$$

(ii) For all non-support vectors, $\beta_i \neq 0$ and $y_i - \overline{W}^{*T} \cdot \overline{x}_i \neq 0$.

hence $\beta_i \neq 0$ for non support vectors as

$$\beta_i \neq 0 \Rightarrow \xi_i \neq 0 \Rightarrow y_i - \overline{W}^{*T} \overline{x}_i \neq 0.$$

All non-support vectors don't lie on ~~margin plan~~ obtained regression line ($\overline{W}^*$).

(c) Dis advantage over SVM

In SVM, for support vectors, $\alpha_i^* \neq 0$.

and in regression, for non-support vectors, $\beta_i^* \neq 0$

The calculation of $\overline{W}^*$ depends on $\overline{\alpha}^*$ and $\overline{\beta}^*$ for both cases respectively.

Since, no. of support vectors are less as compared to no. of non-support vectors in both the cases, $\overline{\alpha}^*$ (of SVM) will have many zeros as its components and $\overline{\beta}^*$ (of regression) will have few zeros as its component.

Hence, calculating $\overline{W}^*$ for SVM is computationally efficient in case of SVM as compared to our regression optimization problem.