# Assignment - 2

## CS771A

Deepak Kumar
(12228)

### Ans 3

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

To show $e^{K(\bar{x},\bar{z})}$ is also kernel.

Doing taylor expansion of $e^{K(\bar{x},\bar{z})}$ gives

$$\cancel{Kt}\; e^{K(\bar{x},\bar{z})} = 1 + \frac{K_1(\bar{x},\bar{z})}{1!} + \frac{K_1^2(\bar{x},\bar{z})}{2!} + \frac{K_1^3(\bar{x},\bar{z})}{3!} + \cdots$$

But we know from class slides that

$$\cancel{Kt}\;\text{are all Kernels}\begin{cases} K(\bar{x},\bar{z}) = K_2(\bar{x},\bar{z}) + C \\ K(\bar{x},\bar{z}) = K_2(\bar{x},\bar{z}) \cdot K_3(\bar{x},\bar{z}) \\ K(\bar{x},\bar{z}) = \alpha K_2(\bar{x},\bar{z}) + \beta K_3(\bar{x},\bar{z}) \end{cases} \qquad \alpha,\beta,C \in \mathbb{R}^+$$

In the above problem $C = 1$, $\alpha,\beta \approx \frac{1}{1!}, \frac{1}{2!} \cdots$

Hence, $K(\bar{x},\bar{z})$ is also a kernel

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

$$K_1(\bar{x},\bar{z}) = e^{\|\bar{x}\|^2} e^{\|\bar{z}\|^2} = g(\bar{x}) \cdot g(\bar{z})$$

But $g: \mathcal{X} \to \mathbb{R}$ and hence $K_1(\bar{x},\bar{z})$ is a kernel

$$K_2(\bar{x},\bar{z}) = \left\langle \frac{\bar{x}}{\|\bar{x}\|^2}, \frac{\bar{z}}{\|\bar{z}\|^2} \right\rangle$$ $\Rightarrow$ Can be written in the form of inner product

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

Since product of kernels is also kernel

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

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

By Mercer's theorem (taught in class) we know that, if we can find a mapping $\phi$ such that $K(\bar{x}, \bar{z}) = \phi(\bar{x}) \cdot \phi(\bar{z})$ then K is psd.

We define a function $\phi$ as follows:

$$\phi(\bar{x}_i) = [\underbrace{1,1,1,\ldots 1}_{x_1 \text{ times}}, \underbrace{0,0,0,\ldots 0}_{N-x_1 \text{ times}}, \quad N \equiv \max_i(\|x_i\|, \|z_i\|)$$

$$\underbrace{1,1,\ldots 1}_{x_2 \text{ times}}, \underbrace{0,0,\ldots\ldots 0}_{N-x_2 \text{ times}}$$

$$\vdots$$

$$\underbrace{1,1,\ldots 1}_{x_d \text{ times}}, \underbrace{0,0,\ldots 0}_{N-x_d \text{ times}}]$$

Now if we take dot product of $\phi(\bar{x})$ & $\phi(\bar{z})$

we get the $\min\{x_i, z_i\}$ component (Equal to 1 & rest 0)

$$\Rightarrow K(\bar{x}, \bar{z}) = \sum_{i=1}^{d} \min\{x_i, z_i\} = \phi(\bar{x}) \cdot \phi(\bar{z})$$

$\Rightarrow K(\bar{x}, \bar{z})$ is kernel by Mercer's Theorem.

Ans 4 :

(a) Given optimisation problem:

$$\underset{\overline{\omega}, \overline{\varepsilon}}{\text{minimise}} \quad L = \sum_{i=1}^{n} \varepsilon_i^2$$

subject to $y_i - \overline{\omega}^T \overline{x} = \varepsilon_i \quad \forall i = 1, 2, 3, \dots n$

$\|w\|_2 \leq B$ (Regularisation parameter)

$$L(\overline{\omega}, \overline{\varepsilon}, \overline{\alpha}, \overline{\mu}) = \sum_{i=1}^{n} \varepsilon_i^2 + \overline{\alpha}(\|w\|_2^2 - B^2) + \sum_{i=1}^{n} \mu_i (y_i - \overline{\omega}^T \overline{x} - \varepsilon_i)$$

$$\frac{\partial L}{\partial \overline{\omega}} = 2\overline{\alpha}\overline{\omega} + \sum_{i=1}^{n} \mu_i(-\overline{x}) = 0$$

$$\Rightarrow \overline{\omega} = \frac{1}{2\|\overline{\alpha}\|} \sum_{i=1}^{n} \mu_i x_i \quad \text{(Dropping the vector condition)}$$

$$\frac{\partial L}{\partial \varepsilon_i} = 2\varepsilon_i - \mu_i = 0 \quad \Rightarrow \varepsilon = \frac{1}{2}\mu_i$$

$$f(\overline{\alpha}, \overline{\mu}) = \min_{\overline{\omega}, \overline{\varepsilon}} L(\overline{\omega}, \overline{\varepsilon}, \overline{\alpha}, \overline{\mu})$$

$$= \frac{\overline{\mu}^T \overline{\mu}}{4} + \overline{\alpha}\left[\frac{1}{4\|\overline{\alpha}\|} \sum_{i=1}^{n}\sum_{j=1}^{n} \mu_i \mu_j \langle x_i, x_j \rangle - B^2\right] + \left[\sum_{i=1}^{n} \mu_i \left(y_i - \frac{1}{2\|\overline{\alpha}\|} \sum_{j=1}^{n} \mu_j \langle x_j, x_i \rangle - \frac{\mu_i}{2}\right)\right]$$

<u>Dual Problem</u>: $f(\overline{\alpha}, \overline{\mu}) = -\frac{\overline{\mu}^T \overline{\mu}}{4} + \sum_{i=1}^{n} \mu_i y_i - \frac{\sum_i \sum_j \mu_i \mu_j \langle x_i, x_j \rangle}{4\|\overline{\alpha}\|} - \overline{\alpha} B^2$

Solving the dual problem to get the max value &

$$\max(f(\overline{\alpha}, \overline{\mu})) = \min\left[\frac{\overline{\mu}^T \mu}{4} - \overline{\mu}^T y + \frac{\overline{\mu}^T \overline{x}^T \overline{x} \, \overline{\mu}}{4\|\overline{\alpha}\|} + \overline{\alpha} B^2\right]$$

$$\frac{\partial f}{\partial \bar{\mu}} = 0 \Rightarrow \frac{\bar{\mu}}{2} - y + \frac{\bar{x}^T \bar{x} \bar{\mu}}{2\|\bar{\alpha}\|} = 0 \Rightarrow \bar{\mu} = y\left[\frac{I}{2} + \frac{\bar{x}^T\bar{x}}{2\|\bar{\alpha}\|}\right]^{-1}$$

$$\frac{\partial f}{\partial \bar{\alpha}} = 0 \Rightarrow \frac{\bar{\mu}^T \bar{x}^T \bar{x} \bar{\mu}}{4\|\bar{\alpha}\|^2} = B^2 \Rightarrow \bar{\mu}^T \bar{x}^T \bar{x} \bar{\mu} = 4\|\bar{\alpha}\|^2 B$$

Hence from previous result of KKT Theorem:

$$\bar{\omega} = \frac{\sum_{i=1}^{n} \mu_i x_i}{2\|\bar{\alpha}\|} = \frac{\bar{x}^T \bar{\mu}}{2\|\bar{\alpha}\|}$$

$$= \frac{\bar{x}^T y}{2\|\bar{\alpha}\|}\left[\frac{I}{2} + \frac{\bar{x}^T\bar{x}}{2\|\bar{\alpha}\|}\right]^{-1}$$

$$\Rightarrow \bar{\omega} = \bar{x}^T\left[\|\bar{\alpha}\| I + \bar{x}^T\bar{x}\right]^{-1} y$$

$$\Rightarrow \bar{\omega}^* = \frac{\sum \mu_i^* x_i}{2\|\bar{\alpha}\|}$$

(b) From above we can see:

$$\bar{\omega}^* = \frac{\sum \mu_i^* x_i}{2\|\bar{\alpha}\|}$$ → Similar to the above SVM formulation.

(c) One of the disadvantage is that $\bar{\omega}$ is dependent on slack variables $(\varepsilon_i)$

$$\because \mu_i \neq 0 \Rightarrow \mu_i(y_i - \bar{\omega}^T x_i - \varepsilon_i) = 0 \Rightarrow y_i - \omega^T x_i = \varepsilon_i$$

Another is summation over the entire training set for a given test vector whereas in SVM we need just sum over the support vectors.