A3) (a) If K_1 is a kernel on $\mathcal{S}$, then $K(\vec{x},\vec{z}) = e^{K_1(\vec{x},\vec{z})}$ is a kernel

K_1 is a ~~kernel fn~~ psd $\Rightarrow K_2^{(1)} = K_1 \circ K_1$ is a psd ~~kernel fn~~ (Hadamard Prod.)

(Schur's Thm.)

$\Rightarrow K^{(2)} = K_1 \circ K'$ is a kernel fn.

In general let $K^{(n)} = \underbrace{K_1 \circ K_1 \ldots \circ K_1}_{n \text{ times.}}$

$\Rightarrow K^{(n)}$, where, $K^{(n)}_{i,j} = (K_{1_{i,j}})^n$, is p.s.d.

$K^{(n)}$ is p.s.d. $\forall n \geq 1$

$$\Rightarrow \vec{x}^T K^{(n)} \vec{x} = \sum_i \sum_j x_i x_j K^{(n)}_{i,j} \geq 0$$

$$= \sum_i \sum_j x_i x_j \frac{K^{(n)}_{i,j}}{n!} \geq 0.$$

Now, summing over n gives,

$$\sum_i \sum_j x_i x_j \sum_{n=1}^{\infty} \frac{K^{(n)}_{i,j}}{n!} \geq 0$$

$$\Rightarrow \sum_i \sum_j x_i x_j \, e^{K_{1_{i,j}}} \geq 0$$

$\therefore e^{K_1(\vec{x},\vec{z})}$ is a kernel fn.

Scanned by CamScanner

(b) ~~I~~ Take fn. ϕ to be

$$\phi(\vec{x}) = \frac{e^{||\vec{x}||^2}}{||\vec{x}||^2} \cdot \vec{x}$$

$$\langle \phi(\vec{x}), \phi(\vec{y}) \rangle = \left\langle \frac{e^{||\vec{x}||^2}}{||\vec{x}||^2} \cdot \vec{x}, \; \frac{e^{||\vec{y}||^2}}{||\vec{y}||^2} \cdot \vec{y} \right\rangle$$

$$= \frac{e^{||\vec{x}||^2 + ||\vec{y}||^2}}{||\vec{x}||^2 \cdot ||\vec{y}||^2} \vec{x}^T \vec{y}$$

$$= K(x, y)$$

(c)

A1) a) The problem is

$$\min_{w, \xi_i} \quad L = \sum_{i=1}^{n} \xi_i^2 = \sum_{i=1}^{n} (y_i - w^T x_i)^2$$

subject to $\|w\|_2 \leq B$

The dual is,

maximize$_\alpha$ $\quad L = \sum (y_i - w^T x_i)^2 + \alpha (\|w\|_2^2 - B^2)$

where $\alpha \geq 0$.

Now, we apply KKT conditions!

$$\frac{\partial}{\partial w}(L) = 0$$

$$X^T(Xw - y) = \alpha w$$

solving, we get

$$\vec{w} = (X^T X + \alpha I)^{-1} X^T y$$

where $X_{n \times n} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$

and $Y_{n \times 1} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$

A few vals. of α may be forbidden so that $(X^T X + \alpha I)^{-1}$ exists. Also $\vec{w}$ is a fn. of α

(b) Yes. Indeed the problem has an equivalent of support vectors. Note that not all points affect the ~~state~~ final val of $\vec{w}$. Therefore, the ones that do, form the equivalent of support vectors.

(c) The dependence of $\vec{w}$ on t is a problem. This did not occur in case of SVM's.