2)

(a) During the kernel trick we replace the the "inner" dot product in the SVM formula (ie).

we optimize

$$\max \left\{ \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \underbrace{\boxed{x_i^T x_j}}_{K(x_i, x_j)} \right.$$

The value $x_i^T x_j$ is nothing but the dot product

$\therefore x_i^T x_j = x_j^T x_i$ } Dot product is commutative

$\therefore K(x_i, x_j) = K(x_j, x_i)$

To test the theory, let us assume a

(i) a polynomial kernel

$$K(x_i, x_j) = (1 + x_i \cdot x_j)^2$$
$$= 1 + (x_i x_j)^2 + 2 x_i x_j$$

$$K(x_j, x_i) = 1 + (x_j x_i)^2 + 2 x_j x_i$$

hence $K(x_i, x_j) = \{\phi(x_i), \phi(x_j)\}$
$= \{\phi(x_j), \phi(x_i)\} = K(x_j, x_i)$

(ii) A Gaussian Kernel.

$$K(x_i, x_j) = e^{\frac{-\|x_i - x_j\|^2}{2\sigma^2}}$$

$$K(x_j, x_i) = e^{\frac{-\|x_j - x_i\|^2}{2\sigma^2}}$$

Since $\|x_j - x_i\|^2 = \|x_j - x_i\|^2$ } Euclidean distance.

we have $K(x_i, x_j) = K(x_j, x_i)$

Hence, $\therefore$ $\boxed{\begin{aligned} K(x_i, x_j) &= (\phi(x_i), \phi(x_j)) \\ &= (\phi(x_j), \phi(x_i)) = K(x_j, x_i) \end{aligned}}$

b) $\|\phi(x_i) - \phi(x_j)\|^2$

Expanding the square

$$= (\phi(x_i), \phi(x_i)) + (\phi(x_j), \phi(x_j)) - 2 \times (\phi(x_i), \phi(x_j))$$

$$= K(x_i, x_i) + K(x_j, x_j) - 2 \cdot K(x_i, x_j)$$

When we use a Gaussian Kernel,

$$K(x_i, x_i) = e^{-\left[\frac{(x_i - x_j)^2}{2}\right]} = e^0 = 1$$

$$\therefore \|\phi(x_i) - \phi(x_j)\|^2 = 1 + 1 - 2\,[\text{some +ve value}]$$

$$= 2 - 2\,(\text{+ve value})$$

This can never be greater than 2

Hence $\|\phi(x_i) - \phi(x_j)\|^2 < 2$ — (1)

Equality occurs when x_i & x_j are very far apart (ie) Euclidian distance is large

$$\therefore e^{-\infty} = 0$$

$$\therefore \|\phi(x_i) - \phi(x_j)\|^2 = 2 \quad \text{--- (2)}$$

Hence from (1) & (2)

$$\boxed{\|\phi(x_i) - \phi(x_j)\|^2 \leq 2}$$