1) for calculating the covariance in higher dimension

i.e. $x_i \rightarrow \phi(x)$ the data $\phi(x_i)$ may not be zero mean

So the higher dimensional dataset should be modified as zero mean data

i.e. $\tilde{\phi}(x_i) = \phi(x_i) - \frac{1}{n}\sum_{k=1}^{n}\phi(x_k)$ ⇒ centered feauters

the corresponding kernel is

$$\tilde{K}_{ij} = \tilde{\phi}(x_i)^T \tilde{\phi}(x_j)$$

$$= \left[\phi(x_i) - \frac{1}{n}\sum_{k=1}^{n}\phi(x_k)\right]^T \left[\phi(x_j) - \frac{1}{n}\sum_{k=1}^{n}\phi(x_k)\right]$$

$$= \phi(x_i)\cdot\phi(x_j) - \frac{1}{n}\sum_{k=1}^{n}\phi(x_i)^T\phi(x_k) - \frac{1}{n}\sum_{k=1}^{n}\phi(x_k)^T\phi(x_j)$$
$$+ \frac{1}{n^2}\sum_{k=1}^{n}\cancel{\phi(x_k)}\sum_{p=1}^{n}\phi(x_k)\cdot\phi(x_p)$$

$$= k(x_i, x_j) - \frac{1}{n}\sum_{k=1}^{n}k(x_i, x_k) - \frac{1}{n}\sum_{k=1}^{n}k(x_j, x_k)$$
$$+ \frac{1}{n^2}\sum_{k,p=1}^{n}\cancel{\phi}\,k(x_k, x_p)$$

writing this in matrix form we get

$$\tilde{k} = k - 2\,1_{1/n}k + 1_{1/n}k\,1_{1/n}$$

where $1_{1/n}$ is a matrix with all elements $1/n$

This is useful because we can operate in lower dimension to center the data.

3) Representer theorem:-

$$J(f) = \underset{f \in \mathcal{F}}{\arg\min} \left[L_y\left(f(x_1) \cdots f(x_n)\right) + \Omega\left(\|f\|_H^2\right) \right]$$

where (x_1, y_1) (x_2, y_2) $\cdots$ (x_n, y_n) are set of observations L is a loss function depends on $f(x_i)$ and Ω is nondecreasing

then the solution takes the form $f^* = \sum_{i=1}^{n} \alpha_i k(x_i, \cdot)$

Proof:-

Denote $f_{\parallel}$ projection of f onto the subspace $\operatorname{span}\{k(x_i, \cdot);\ 1 \leq i \leq n\}$

Such that $f = f_{\parallel} + f_{\perp}$

where $f_{\parallel} = \sum_{i=1}^{n} \alpha_i k(x_i, \cdot)$

Regularize $\|f\|_H^2 = \|f_{\parallel}\|_H^2 + \|f_{\perp}\|_H^2 \geq \|f\|_H^2$

$$\Omega\left(\|f\|_H^2\right) \geq \Omega\left(\|f_{\parallel}\|_H^2\right)$$

so this form is minimized for $f = f_{\parallel}$

Individual terms $f(x_i)$ in the loss

$f(x_i) = \langle f, k(x_i, \cdot)\rangle_H = \langle f_{\parallel} + f_{\perp}, k(x_i, \cdot)\rangle_H$

$= \langle f_{\parallel}, k(x_i, \cdot)\rangle + 0 \quad (\because \langle f_{\perp}, k(x_i, \cdot)\rangle = 0$

$\because$ orthogonal to the span of vectors)

$\therefore L_y(f(x_1) \dots f(x_n)) = L_y(f_{\parallel}(x_1) \dots f_{\parallel}(x_n))$

=) Here $L(\cdot)$ depends on the component of f in data subspace i.e. $f_{\parallel}$

Regularizer $\Omega(\cdot)$ minimized when $f = f_{\parallel}$ } ~~Hence theorem.~~

$\therefore$ optimized solution $f = \sum_{i=1}^{n} \alpha_i k(x_i, \cdot)$

3) Application of Representer theorem in solving Ridge Regression.

$$f^* = \underset{f \in F}{\arg\min} \; \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2 + \lambda \|f\|_H^2$$

$$f(\cdot) = \sum_{i=1}^{n} \alpha_i k(\cdot, x_i)$$

$$f^* = \underset{f \in F}{\arg\min} \; \frac{1}{n} \sum_{i=1}^{n} \Big(y_i - \sum_{j=1}^{n} \alpha_j k(x_i, x_j)\Big)^2 + \lambda \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j k(x_i, x_j)$$

converting into Matrix form

$$f^* = \underset{\alpha \in \mathbb{R}^n}{\arg\min} \; \frac{1}{n} \|y - K\alpha\|^2 + \lambda \alpha^T K \alpha$$

$J(\alpha) = \frac{1}{n}(\underline{y} - K\underline{\alpha})^T(\underline{y} - K\underline{\alpha}) + \lambda\underline{\alpha}^T K\underline{\alpha}$ is a sum of two

convex functions of α

$$\alpha^* = \underset{\alpha \in \mathbb{R}^n}{\arg\min}\ J(\underline{\alpha})$$

$$\nabla J(\underline{\alpha}) = 0 \Rightarrow \nabla\left[\frac{(\underline{y} - K\underline{\alpha})^T(\underline{y} - K\underline{\alpha})}{n} + \lambda\underline{\alpha}^T K\underline{\alpha}\right] = 0$$

$$\therefore \frac{-1}{n}\left(2K^T(\underline{y} - K\underline{\alpha})\right) + 2\lambda K\underline{\alpha} = 0$$

$$\frac{-1}{n}\left[\underline{y} - K\underline{\alpha}\right] + \lambda\underline{\alpha} = 0$$

$$\underline{\alpha} = [K + n\lambda I]^{-1}\underline{y}$$

for a datapoint x the optimized y is

$$\hat{y} = \underline{\alpha}^T K(\underline{x})$$

where $K(\underline{x}) = [k(x_1, x) \;\cdots\; k(x_n, x)]^T$

4) Representer theorem to solve kernel S.V.M.

$$f^* = \underset{f \in F}{\arg\min}\ \max(1 - y_i f(x_i), 0) + \lambda\|f\|_H^2$$

$$\Rightarrow \min \lambda\|f\|_H^2$$

sub to $y_i f(x_i) - 1 \geq 0$

let $f(\cdot) = \sum_i r_i k(x_i, \cdot)$

$\Rightarrow \min \lambda \sum_j \sum_i r_i r_j k(x_i, x_j)$

sub. to. $y_i \sum_{j=1}^{n} r_j k(x_i, x_j) - 1 \geq 0$

let $\lambda = 1/2$

~~$\Rightarrow$ writting in matrix~~

$\Rightarrow$ writing in Matrix form, we get

$\frac{1}{2} R^T K R + \left[\sum_{i=1}^{n} \alpha_i - \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i y_i r_j K(x_i, x_j)\right]$ $\Rightarrow$ legrange Multiplier

$\Lambda = \begin{bmatrix} \alpha_1 y_1 \\ \vdots \\ \alpha_n y_n \end{bmatrix}$ $R = \begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix}$

$\Rightarrow \frac{1}{2} R^T K R - \Lambda K R + \sum_{i=1}^{n} \alpha_i$

Diff wrt. $R \Rightarrow RK - \Lambda K = 0$

$R = \Lambda$

$r_i = \alpha_i y_i$

Substituting back we get

$L_D = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j)$

s.t. $\alpha_i \geq 0$