1 (a)

$$K^c_{ij} = K_c(x_i, x_j) = \langle \phi_c(x_i), \phi_c(x_j) \rangle$$

$$\phi_c(x) = \phi(x) - \frac{1}{N}\sum_{i=1}^{N}\phi(x_i)$$

$$\therefore = \langle \phi_c(x_i)\ \phi_c(x_j) \rangle = \left[\phi(x_i) - \frac{1}{N}\sum_{p=1}^{N}\phi(x_p)\right]^T\left[\phi(x_j) - \frac{1}{N}\sum_{q=1}^{N}\phi(x_q)\right] - (1)$$

$$= \phi^T(x_i)\phi(x_j) - \frac{1}{N}\left\{\sum_{p=1}^{N}\phi^T(x_p)\right\}\phi(x_j) +$$

$$-\frac{1}{N}\phi^T(x_i)\sum_{q=1}^{N}\phi(x_q) + \frac{1}{N^2}\left\{\sum_{p=1}^{N}\phi^T(x_p)\right\}\left\{\sum_{q=1}^{N}\phi(x_q)\right\} \quad - (2)$$

$$= K_{ij} - \frac{1}{N}\sum_{p=1}^{N}\phi^T(x_p)\cdot\phi(x_j) - \frac{1}{N}\sum_{q=1}^{N}\phi^T(x_i)\cdot\phi(x_q)$$

$$+ \frac{1}{N^2}\sum_{p=1}^{N}\sum_{q=1}^{N}\phi^T(x_p)\cdot\phi(x_q) \quad - (3)$$

$$= K_{ij} - \frac{1}{N}\sum_{p=1}^{N}K_{pj} - \frac{1}{N}\sum_{q=1}^{N}K_{iq} + \frac{1}{N^2}\sum_{p=1}^{N}\sum_{q=1}^{N}K_{pq} \quad - (4)$$

Here, in Eq 3. $\frac{1}{N}\sum_{p=1}^{N}\phi^T(x_p)\cdot\phi(x_j)$ for each value of p. $\phi^T(x_p)$ is a vector, of which, we find a dot product with. $\phi(x_j)$. if we have only 1 "x_p" then we have $\phi^T(x_p)\cdot\phi(x_j) = K_{pq}$. but we have $p \in \{1, \dots, N\}$ which we are averaging over. that's why we have. $\frac{1}{N}\sum_{p=1}^{N}K_{pj}$.

1 (b)

b. $K_c(x_i, t_j) = \langle \phi_c(x_i), \phi_c(t_j) \rangle$ We need to center the data w.r.t the training dataset.

$$= \left(\phi(x_i) - \frac{1}{N}\sum_{p=1}^{N}\phi(x_p)\right)^T\left(\phi(t_j) - \frac{1}{N}\sum_{q=1}^{N}\phi(x_q)\right)$$

$$= \left[\phi^T(x_i) - \frac{1}{N}\sum_{p=1}^{N}\phi^T(x_p)\right]\left[\phi(t_j) - \frac{1}{N}\sum_{q=1}^{N}\phi(x_q)\right]$$

$$= \phi^T(x_i)\phi(t_j) - \frac{1}{N}\phi^T(x_i)\sum_{q=1}\phi(x_q) - \frac{1}{N}\left\{\sum_{p=1}\phi^T(x_p)\right\}\phi(t_j)$$

$$+ \frac{1}{N^2}\sum_{p=1}^{N}\phi^T(x_p)\cdot\sum_{q=1}^{N}\phi(x_q)$$

$$= K(x_i, t_j) - \frac{1}{N}\sum_{q=1}^{N}\phi^T(x_i)\phi(x_q) - \frac{1}{N}\sum_{p=1}^{N}\phi^T(x_p)\cdot\phi(t_j) + \frac{1}{N^2}\sum_{p=1}^{N}\sum_{q=1}^{N}\phi^T(x_p)\phi(x_q)$$

$$= \underbrace{K(x_i, t_j)}_{\text{from (train, test) Kernel}} - \underbrace{\frac{1}{N}\sum_{q=1}^{N}K_{iq}}_{\text{Terms from (Train, Train) Kernel}} - \underbrace{\frac{1}{N}\sum_{p=1}^{N}K(x_p, t_j)}_{\text{from (Train, Test) Kernel}} + \underbrace{\frac{1}{N^2}\sum_{p=1}^{N}\sum_{q=1}^{N}K_{pq}}_{\text{These terms Come from the (Train, Train) Kernel}}$$

2 (b)

(b) Derivation of dual problem.

Expressing the primal problem in its canonical form.

$$\min_{\omega, \xi_\varsigma} \; \frac{1}{2}\|\omega\|^2 + \frac{c}{2}\sum_{i=1}^{m} \xi_{\varsigma i}^2.$$

$$\text{s.t.} \quad 1 - (y_i(\omega^T x_i)) - \xi_{\varsigma i} \leq 0$$

NOTE: here "b" is included as a column of constants in the vector 'x'.

let us define a lagrange variable λ and derive the Lagrangian dua

i.e.

$$\min_{\omega, \xi_\varsigma} \; \frac{1}{2}\|\omega\|^2 + \frac{c}{2}\sum_{i=1}^{m} \xi_{\varsigma i}^2 + \sum_{i=1}^{m}\lambda_i\left(1 - y_i(\omega^T x_i) - \xi_{\varsigma i}\right) \quad \forall \; i = 1 \ldots m. \quad - (1)$$

$$= L(\omega, \xi_\varsigma, \lambda) \qquad \text{s.t.} \; \lambda_i \geq 0.$$

· Finding the minimal of the lagrangian by finding the derivative w.r.t the primal variables and setting the derivative to zero.

$$\nabla_\omega L(\omega, \xi_\varsigma, \lambda) = \omega - \sum_{i=1}^{m}\lambda_i y_i x_i = 0.$$

$$\therefore \; \omega = \sum_{i=1}^{m}\lambda_i y_i x_i \quad - (2).$$

$$\nabla_{\xi_{\varsigma i}} L(\omega, \xi_\varsigma, \lambda) = C\xi_{\varsigma i} - \lambda_i = 0 \qquad \therefore \; \xi_{\varsigma i} = \frac{\lambda_i}{C} \quad - (3)$$

$$\text{or} \quad C = \frac{\lambda_i}{\xi_{\varsigma i}}$$

Substituting (2) and (3) in (1) ~~we~~ we get:

$$\cancel{\min} \; \frac{1}{2}\left\|\sum_{i=1}^{m}\lambda_i y_i x_i\right\|^2 + \cancel{\frac{C}{2}} \sum_{i=1}^{m}\frac{\lambda_i}{\xi_{\varsigma i}} \times \frac{\xi_{\varsigma i}^2}{2} + \sum_{i=1}^{m}\lambda_i\left(1 - y_i\left(\sum_{j=1}^{m}\lambda_j y_j x_j\right)^T x_i - \right.$$

$$= \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j x_i^T x_j + \sum_{i=1}^{m}\frac{\lambda_i \xi_{\varsigma i}}{2} + \sum_{i=1}^{m}\lambda_i - \sum_{i=1}^{m}\lambda_i\xi_{\varsigma i} - \sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j x_j^T x_i$$

$$= \sum_{i=1}^{m}\lambda_i - \sum_{i=1}^{m}\frac{\lambda_i\xi_{\varsigma i}}{2} - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j \underbrace{\langle x_i, x_j\rangle}_{\text{Can be replaced by a Kernel}} \; ; \text{s.t } \lambda_i$$

$$= \max_{\lambda_\varsigma} \sum_{i=1}^{m}\lambda_i\left(1 - \frac{\xi_{\varsigma i}}{2}\right) - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j y_i y_j K(x_i, x_j)$$

(The dual problem)

$$\lambda_i \geq 0$$

2 (a)

a). By definition, the constraint $C > 0$ controls the balance between maximizing the margin & amount of slack needed. also $\lambda_i \geq 0$ $\therefore$ from Eq3.

$$C = \frac{\lambda_i}{\varepsilon_{si}} \quad \therefore \; \varepsilon_{si} \geq 0$$

$$\varepsilon_i = \frac{\lambda_i}{c}$$

$$\Rightarrow \frac{\lambda_i \geq 0}{c > 0} \quad \therefore \; \varepsilon_i \geq 0$$

2 (c)

c). In this problem, let α be the dual variable, the complementary slackness conditions are $\alpha_i \left[y_i(\omega^T x_i) - 1 + \varepsilon_{si} \right] = 0$ for $i = 1 \dots m$.

what is the value (or range) of the margin $y_i(\omega^T x_i + b)$ for $\alpha_i > 0$ and $\alpha_i = 0$.

at the optimal condition. ($\overset{\omega =}{\omega^*}$). the complementary slackness condition holds.

i.e. if $\alpha_i > 0$ then $y_i [(\omega^{*T} x_i) - 1 + \varepsilon_{si}] = 0$.

$$\boxed{y_i (\omega^{*T} x_i) = 1 - \varepsilon_{si}}$$

and when $\alpha_i = 0$ then $y_i (\omega^{*T} x_i) - 1 + \varepsilon_{si} < 0$

$$\boxed{y_i (\omega^{*T} x_i) < 1 - \varepsilon_{si}}$$