

1 (a)

$$\begin{aligned}
 K_{ij}^c &= K_c(x_i, x_j) = \langle \phi_c(x_i), \phi_c(x_j) \rangle \\
 \phi_c(x) &= \phi(x) - \frac{1}{N} \sum_{i=1}^N \phi(x_i) \\
 \therefore \langle \phi_c(x_i), \phi_c(x_j) \rangle &= \left[ \phi(x_i) - \frac{1}{N} \sum_{p=1}^N \phi(x_p) \right]^T \left[ \phi(x_j) - \frac{1}{N} \sum_{q=1}^N \phi(x_q) \right] \quad \text{--- (1)} \\
 &= \phi^T(x_i) \phi(x_j) - \frac{1}{N} \left\{ \sum_{p=1}^N \phi^T(x_p) \right\} \phi(x_j) \\
 &\quad - \frac{1}{N} \phi^T(x_i) \sum_{q=1}^N \phi(x_q) + \frac{1}{N^2} \left\{ \sum_{p=1}^N \phi^T(x_p) \right\} \left\{ \sum_{q=1}^N \phi(x_q) \right\} \quad \text{--- (2)} \\
 &= K_{ij} - \frac{1}{N} \sum_{p=1}^N \phi^T(x_p) \cdot \phi(x_j) - \frac{1}{N} \sum_{q=1}^N \phi^T(x_i) \cdot \phi(x_q) \\
 &\quad + \frac{1}{N^2} \sum_{p=1}^N \sum_{q=1}^N \phi^T(x_p) \cdot \phi(x_q) \quad \text{--- (3)} \\
 &= K_{ij} - \frac{1}{N} \sum_{p=1}^N K_{pj} - \frac{1}{N} \sum_{q=1}^N K_{iq} + \frac{1}{N^2} \sum_{p=1}^N \sum_{q=1}^N K_{pq} \quad \text{--- (4)}
 \end{aligned}$$

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  $\phi^T(x_p) \cdot \phi(x_j) = K_{pj}$  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.

$$\begin{aligned}
 &= \left( \phi_c(x_i) - \frac{1}{N} \sum_{p=1}^N \phi(x_p) \right)^T \left( \phi_c(t_j) - \frac{1}{N} \sum_{q=1}^N \phi(x_q) \right) \\
 &= \left[ \phi^T(x_i) - \frac{1}{N} \sum_{p=1}^N \phi^T(x_p) \right] \left[ \phi(t_j) - \frac{1}{N} \sum_{q=1}^N \phi(x_q) \right] \\
 &= \phi^T(x_i) \phi(t_j) - \frac{1}{N} \phi^T(x_i) \sum_{q=1}^N \phi(x_q) - \frac{1}{N} \left\{ \sum_{p=1}^N \phi^T(x_p) \right\} \phi(t_j) \\
 &\quad + \frac{1}{N^2} \sum_{p=1}^N \phi^T(x_p) \sum_{q=1}^N \phi(x_q) \\
 &= K(x_i, t_j) - \frac{1}{N} \sum_{q=1}^N \phi^T(x_i) \phi(x_q) - \frac{1}{N} \sum_{p=1}^N \phi^T(x_p) \cdot \phi(t_j) + \frac{1}{N^2} \sum_{p=1}^N \sum_{q=1}^N \phi^T(x_p) \phi(x_q) \\
 &= \underbrace{K(x_i, t_j)}_{\text{from (train, test) Kernel}} - \underbrace{\frac{1}{N} \sum_{q=1}^N K_{iq}}_{\text{Terms from (Train, Train) Kernel}} - \underbrace{\frac{1}{N} \sum_{p=1}^N K(x_p, t_j)}_{\text{from (Train, Text) Kernel}} + \underbrace{\frac{1}{N^2} \sum_{p=1}^N \sum_{q=1}^N K_{pq}}_{\text{These terms come from the (Train, Train) Kernel}}
 \end{aligned}$$

2 (b)

(b) Derivation of dual problem.

Expressing the primal problem in its canonical form.

$$\min_{w, \xi_i} \quad \frac{1}{2} \|w\|^2 + \frac{c}{2} \sum_{i=1}^m \xi_i^2$$

$$\text{s.t.} \quad 1 - (y_i (w^T x_i)) - \xi_i \leq 0$$

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

let us define a lagrange variable  $\lambda$  and derive the Lagrangian dual

$$\text{i.e.} \quad \min_{w, \xi_i} \quad \frac{1}{2} \|w\|^2 + \frac{c}{2} \sum_{i=1}^m \xi_i^2 + \sum_{i=1}^m \lambda_i (1 - y_i (w^T x_i) - \xi_i) \quad \forall i=1 \dots m.$$

$$= L(w, \xi_i, \lambda) \quad \text{s.t.} \quad \lambda_i \geq 0. \quad - (1)$$

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

$$\nabla_w L(w, \xi_i, \lambda) = w - \sum_{i=1}^m \lambda_i y_i x_i = 0.$$

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

$$\nabla_{\xi_i} L(w, \xi_i, \lambda) = c \xi_i - \lambda_i = 0 \quad \therefore \xi_i = \frac{\lambda_i}{c} \quad - (3)$$

$$\text{or } c = \frac{\lambda_i}{\xi_i}$$

Substituting (2) and (3) in (1) we get.

$$\min_{\lambda_i} \quad \frac{1}{2} \left\| \sum_{i=1}^m \lambda_i y_i x_i \right\|^2 + \sum_{i=1}^m \frac{\lambda_i}{\xi_i} \times \frac{\xi_i^2}{2} + \sum_{i=1}^m \lambda_i (1 - y_i (\sum_{j=1}^m \lambda_j y_j x_j)^T x_i)$$

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

$$- \sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j y_i y_j x_i^T x_j$$

$$= \sum_{i=1}^m \lambda_i - \sum_{i=1}^m \frac{\lambda_i \xi_i}{2} - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j y_i y_j \langle x_i, x_j \rangle \quad ; \text{ s.t. } \lambda_i$$

Can be replaced by a kernel

$$= \max_{\lambda_i} \sum_{i=1}^m \lambda_i (1 - \frac{\xi_i}{2}) - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \lambda_i \lambda_j y_i y_j K(x_i, x_j). \quad (\text{The dual problem})$$

$$\lambda_i \geq 0$$

## 2 (a)

a). By definition, the constraint  $c > 0$  controls the balance between maximizing the margin & amount of slack needed. also  $\lambda_i > 0 \therefore$  from Eq 3.

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

$$\Rightarrow \xi_i = \frac{\lambda_i}{c}$$

$$\Rightarrow \frac{\lambda_i > 0}{c > 0} \therefore \xi_i > 0$$

So, this constraint holds implicitly. Hence, its not needed to be included among the constraints in the original optimization problem.

## 2 (c)

c). In this problem, let  $\alpha$  be the dual variable, the complementary slackness conditions are  $\alpha_i [y_i(w^T x_i) - 1 + \xi_i] = 0$  for  $i=1 \dots m$ .

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

at the optimal condition. ( $w^*$ ). the complementary slackness condition holds.

i.e. if  $\alpha_i > 0$  then  $y_i(w^{*T} x_i) - 1 + \xi_i = 0$ .

$$y_i(w^{*T} x_i) = 1 - \xi_i$$

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

$$y_i(w^{*T} x_i) < 1 - \xi_i$$