1. (a) For problem (1), the boundary is

$t^{(i)}(w^T x^{(i)}+b) \geq 1-\xi_i$, $\xi_i \geq 0$, it is equivalent

to $\xi_i \geq \max(0, 1-t^{(i)}(w^T x^{(i)}+b))$,

Thus the problem (1) can be written as

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + c\sum_{i=1}^{N}\xi_i,$$

subject to $\xi_i \geq \max(0, 1-t^{(i)}(w^T x^{(i)}+b))$.

Now let's consider the problem from the perspective of final optimal value of w, b, ξ, set their optimal value of this problem is w^*, b^*, ξ^*. Since $c>0$, we must have $\xi_i^* = \max(0, 1-t^{(i)}(w_*^T x^{(i)}+b^*))$, otherwise we can use $\max(0, 1-t^{(i)}(w^T x^{(i)}+b))$ to replace ξ_i^* and make the target function smaller. Therefore, whatever w^*, b^* is, the optimal ξ_i^* can be always represented by $\max(0, 1-t^{(i)}(w^T x^{(i)}+b))$.

Thus the problem is equivalent to:

$$\min_{w,b} \frac{1}{2}\|w\|^2 + c\sum_{i=1}^{N}\max(0, 1-t^{(i)}(w^T x^{(i)}+b))$$

(b) The margin hyperlane $t^{(i)}((w^*)^T x+b^*)=1$ can be written as $t^{(i)}(w^*)^T x + (t^{(i)}b^*-1)=0$, then the distance is

$$|r| = \frac{|t^{(i)}(w^*)^T x^{(i)} + t^{(i)}b^*-1|}{\|t^{(i)}w^*\|}$$

$$= |\xi_i^*| \cdot \frac{1}{\|t^{(i)}w^*\|}$$

Obviously, $|r|$ is proportional to ξ_i^* ($\xi_i^* \geq 0$)

(c) If for all i, $1-t^{(i)}(w^T x^{(i)}+b) \neq 0$, then

$$\nabla_w E(w,b) = w + c\sum_{i=1}^{N} -t^{(i)}x^{(i)} \cdot \mathbb{1}_{\{1-t^{(i)}(w^T x^{(i)}+b)>0\}}$$

$$\nabla_b E(w,b) = c\sum_{i}^{N} -t^{(i)}\mathbb{1}_{\{1-t^{(i)}(w^T x^{(i)}+b)>0\}}$$

If for some i, $1-t^{(i)}(w^T x^{(i)}+b)=0$, & for other j, $1-t^{(j)}(w^T x^{(j)}+b) \neq 0$, then the derivative is undifined at i, the subderivative is:

$$\nabla_W^- E(w,b) = w + C(\sum_i -t^{(i)}x^{(i)} + C(\sum_j -t^{(j)}x^{(j)} 1_{\{1-t^{(j)}(w^Tx^{(j)}+b)>0\}}$$

$$\nabla_W^+ E(w,b) = w + C \sum_j -t^{(j)}x^{(j)} 1_{\{1-t^{(j)}(w^Tx^{(j)}+b)>0\}}$$

$$\nabla_b^- E(w,b) = (C\sum_i -t^{(i)} + C\sum_j -t^{(j)} 1_{\{1-t^{(j)}(w^Tx^{(j)}+b)>0\}}$$

$$\nabla_b^+ E(w,b) = C\sum_j -t^{(j)} 1_{\{1-t^{(j)}(w^Tx^{(j)}+b)>0\}}$$

(e) If for i, $1-t^{(i)}(w^Tx^{(i)}+b) \neq 0$, then

$$\nabla_W E^{(i)}(w,b) = \frac{w}{N} + C(-t^{(i)} \cdot x^{(i)}) \cdot 1_{\{1-t^{(i)}(w^Tx^{(i)}+b)>0\}}$$

$$\nabla_b E^{(i)}(w,b) = -C t^{(i)} 1_{\{1-t^{(i)}(w^Tx^{(i)}+b)>0\}}$$

if for i, $1-t^{(i)}(w^Tx^{(i)}+b)=0$, then

$$\nabla_W^- E^{(i)}(w,b) = \frac{w}{N} + C(-t^{(i)} \cdot x^{(i)})$$

$$\nabla_W^+ E^{(i)}(w,b) = \frac{w}{N}$$

$$\nabla_b^- E^{(i)}(w,b) = -C t^{(i)}$$

$$\nabla_b^+ E^{(i)}(w,b) = 0$$

(h) $L(w,b,\lambda,v) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^N \xi_i + \sum_{i=1}^N \lambda_i(1-\xi_i - t^{(i)}(w^Tx^{(i)}+b)) - \sum_{i=1}^N v_i\xi_i$

Then the dual problem is $\max_{v,\lambda,\lambda_i,v_i\geq 0} \min_{w,b} L(w,b,\lambda,v)$

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^N \lambda_i t^{(i)}x^{(i)} = 0 \Rightarrow w = \sum_{i=1}^N \lambda_i t^{(i)} x^{(i)}$$

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^N \lambda_i t^{(i)} = 0, \quad \frac{\partial L}{\partial \xi_i} = C - \lambda_i - v_i = 0$$

$$\therefore \tilde{L}(\lambda,v) = \frac{1}{2}[\sum_{i=1}^N \lambda_i t^{(i)}x^{(i)}][\sum_{i=1}^N \lambda_i t^{(i)}x^{(i)}] + \sum_{i=1}^N(\lambda_i+v_i)\xi_i - \sum_{i=1}^N v_i\xi_i +$$

$$\sum_{i=1}^N \lambda_i(1-\xi_i - t^{(i)}((\sum_{i=1}^N \lambda_i t^{(i)}x^{(i)})^T x^{(i)} + b)$$

$$= \frac{1}{2}[\sum_{i=1}^N \lambda_i t^{(i)}x^{(i)}]^T[\sum_{i=1}^N \lambda_i t^{(i)}x^{(i)}] + \sum_{i=1}^N \lambda_i - (\sum_{i=1}^N \lambda_i t^{(i)}x^{(i)})^T(\sum_{i=1}^N \lambda_i t^{(i)}x^{(i)})$$

$$= \sum_{i=1}^N \lambda_i - \frac{1}{2}[\sum_{i=1}^N \lambda_i t^{(i)}x^{(i)}]^T[\sum_{i=1}^N \lambda_i t^{(i)}x^{(i)}]$$

Set $k(x,z) = x^Tz \Leftarrow$ $$= \sum_{i=1}^N \lambda_i - \frac{1}{2}\sum_{i=1}^N\sum_{j=1}^N \lambda_i\lambda_j t^{(i)}t^{(j)}x^{(i)T}x^{(j)} = \sum_{i=1}^N \lambda_i - \frac{1}{2}\sum_{i=1}^N\sum_{j=1}^N \lambda_i\lambda_j t^{(i)}t^{(j)}k(x^{(i)},x^{(j)})$$

$\therefore$ The dual problem is : $\min_\lambda \hat{L}(\lambda) = \sum_{i=1}^N \lambda_i - \frac{1}{2}\sum_{i=1}^N\sum_{j=1}^N \lambda_i\lambda_j t^{(i)}t^{(j)}k(x^{(i)},x^{(j)})$

subject to : $0 \leq \lambda_i \leq C$,

$$\sum_{i=1}^N \lambda_i t^{(i)} = 0$$

3. (a) $K(u,v) = (\langle u,v\rangle + 1)^4$

$$= \langle u,v\rangle^4 + 4\langle u,v\rangle^3 + 6\langle u,v\rangle^2 + 4\langle u,v\rangle + 1$$

For $d=3$, $u=(u_1,u_2,u_3)$, $v=(v_1,v_2,v_3)$

For general mode, $\langle u,v\rangle^p = (u_1v_1 + u_2v_2 + u_3v_3)^p = \sum_{j_1,j_2,j_3,\Sigma j=p} \binom{p}{j_1,j_2,j_3}(u_1v_1)^{j_1}(u_2v_2)^{j_2}(u_3v_3)^{j_3}$

$$\therefore \bar{\Phi}_p(x) = [\cdots, \binom{p}{j_1,j_2,j_3}^{\frac{1}{2}}(u_1)^{j_1}(u_2)^{j_2}(u_3)^{j_3}, \cdots] \quad (*)$$

$p=4$, $\bar{\Phi}_4(u) = [u_1^4, u_2^4, u_3^4, 2u_1^3u_2, 2u_1^3u_3, 2u_1u_2^3, 2u_2^3u_3, 2u_1u_3^3,$
$2u_2u_3^3, \sqrt{6}u_1^2u_2^2, \sqrt{6}u_1^2u_3^2, \sqrt{6}u_2^2u_3^2, 2\sqrt{3}u_1^2u_2u_3, 2\sqrt{3}u_1u_2^2u_3, 2\sqrt{3}u_1u_2u_3^2]$

$p=3$, $\bar{\Phi}_3(u) = [u_1^3, u_2^3, u_3^3, \sqrt{3}u_1^2u_2, \sqrt{3}u_1u_2^2, \sqrt{3}u_1^2u_3, \sqrt{3}u_2^2u_3, \sqrt{3}u_1u_3^2, \sqrt{3}u_2u_3^2, \sqrt{6}u_1u_2u_3]$

$p=2$, $\bar{\Phi}_2(u) = [u_1^2, u_2^2, u_3^2, \sqrt{2}u_1u_2, \sqrt{2}u_1u_3, \sqrt{2}u_2u_3]$

$p=1$, $\bar{\Phi}_1(u) = [u_1, u_2, u_3]$

Then for u, $\phi(u) = (\bar{\Phi}_4(u), 2\bar{\Phi}_3(u), \sqrt{6}\bar{\Phi}_2(u), 2\bar{\Phi}_1(u), 1)$

Thus $k(u,v) = \phi(u)^T\phi(v)$

For arbitrary dimension d, based on the conclusion $(*)$ above,

$p=4$, $\bar{\Phi}_4(u) = [u_1^4, \cdots u_d^4, 2u_1^3u_2, \cdots, \sqrt{6}u_1^2u_2^2, \cdots 2\sqrt{3}u_1^2u_2u_3 \cdots]$

$p=3$, $\bar{\Phi}_3(u) = [u_1^3, \cdots u_d^3, \sqrt{3}u_1^2u_2 \cdots, \sqrt{6}u_1u_2u_3, \cdots]$

$p=2$, $\bar{\Phi}_2(u) = [u_1^2, \cdots u_d^2, \sqrt{2}u_1v_2, \cdots]$

$p=1$, $\bar{\Phi}_1(u) = [u_1, \cdots u_d]$.

Then for u, $\phi(u) = (\bar{\Phi}_4(u), 2\bar{\Phi}_3(u), \sqrt{6}\bar{\Phi}_2(u), 2\bar{\Phi}_1(u), 1)$

satisfying $k(u,v) = \phi(u)^T\phi(v)$

(b) From the definition of Gram matrix k, for a positive-definite kernel, k must be PSD, $k_i = \phi^{(i)}(x)\phi^{(i)}(z)$ is a Gram matrix, for $\forall a \in \mathbb{R}^n$, $a^T k_i a \geq 0$

(i) It is kernel. Since k_i is kernel, for $\forall a \in \mathbb{R}^n$,

$a^T k_1 a \geq 0$, $a^T k_2 a \geq 0$ $\Rightarrow$ $a^T k a = a^T(k_1 + k_2)a = a^T k_1 a + a^T k_2 a \geq 0$

Thus k is kernel.

(ii) Not kernel. Set $k_2 = 2k_1$, then for $\forall a \in \mathbb{R}^n$, $a^T k_1 a \geq 0$

while $a^T k a = a^T(k_1 - 2k_1)a = -a^T k_1 a \leq 0$

(iii) It is kernel. For $\forall m \in \mathbb{R}^n$, $m^T k_1 m \geq 0$, then for $a > 0$

$m^T k m = m^T(a k_1)m = a(m^T k_1 m) \geq 0$

(iv) It is kernel. $k(x,z) = k_1(x,z)\, k_2(x,z)$

$$= \sum_i \phi_i^{(1)}(x)\phi_i^{(1)}(z) \cdot \sum_j \phi_j^{(2)}(x)\phi_j^{(2)}(z)$$

$$= \sum_i \sum_j \phi_i^{(1)}(x)\phi_i^{(1)}(z)\phi_j^{(2)}(x)\phi_j^{(2)}(z)$$

$$= \sum_i \sum_j [\phi_i^{(1)}(x)\phi_{(j)}^{(2)}(x)] \cdot [\phi_{(i)}^{(1)}(z)\phi_{(j)}^{(2)}(z)] = \sum_{i,j} \Phi_{i,j}(x)\Phi_{i,j}(z)$$

Thus k can be written as $k(x,z) = \Phi(x)^T \Phi(z)$ $\Rightarrow$ k is knernel.

(v) It is kernel. Let $\varphi(x) = f(x)$, $f: \mathbb{R}^D \to \mathbb{R}$

$\therefore$ $\varphi^T(x) = f^T(x) = f(x) = \varphi(x)$

$\therefore$ $k(x,z) = \varphi(x)\cdot\varphi(z) = \varphi(x)^T\varphi(z)$ $\Rightarrow$ k is kernel.

(vi) It is kernel: $k(x,z) = a_1 k_1(x,z) + \cdots + a_p k_1^p(x,z)$

From (iv), $k_1^p(x,z)$ is kernel, From (iii), $a_p k_1^p(x,z)$ is kernel.

Then From (i), $k(x,z) = \sum_i a_i k^i(x,z)$ is kernel.

(vii) $k(x,z) = \exp\left(-\frac{\|x-z\|^2}{2\sigma^2}\right) = \exp\left(-\frac{x^Tx}{2\sigma^2}\right)\exp\left(\frac{x^Tz}{\sigma^2}\right)\exp\left(-\frac{z^Tz}{2\sigma^2}\right)$

It can be written as $k(x,z) = f(x)\exp\left(\frac{x^Tz}{\sigma^2}\right)f(z)$

By Taylor expansion, $\exp\left(\frac{x^Tz}{\sigma^2}\right) = \sum_i^\infty \frac{\left(\frac{x^Tz}{\sigma^2}\right)^n}{n!}$,

Noticing x^Tz is a kernel, From (vi), $\sum^\infty \left(\frac{x^Tz}{\sigma^2}\right)^n / n!$ is a kernel.

Thus we can write $\exp\left(\frac{x^Tz}{\sigma^2}\right)$ as $\phi(x)^T\phi(z)$, Then

$k(x,z) = f(x)\langle\phi(x), \phi(z)\rangle f(z) = \langle f(x)\phi(x), f(z)\phi(z)\rangle$

4(a)

$$P^{-1} - P^{-1}Q(R^{-1} + SP^{-1}Q)^{-1}SP^{-1}$$

$$= P^{-1}Q(R^{-1} + SP^{-1}Q)^{-1}(R^{-1} + SP^{-1}Q)Q^{-1} - P^{-1}Q(R^{-1} + SP^{-1}Q)^{-1}SP^{-1}$$

$$= P^{-1}Q(R^{-1} + SP^{-1}Q)^{-1}(R^{-1} + SP^{-1}Q - SP^{-1}Q)Q^{-1}$$

$$= P^{-1}Q(R^{-1} + SP^{-1}Q)^{-1}R^{-1}Q^{-1}$$

From the description of the question, we have

$$(P + QRS)^{-1} = P^{-1}Q(R^{-1} + SP^{-1}Q)^{-1}R^{-1}Q^{-1}$$

with $w = (\Phi^T\Phi + \lambda I)^{-1}\Phi^T t$

Let $P = \lambda I$, $Q = \Phi^T$, $R = I$, $S = \Phi$, we have

$$(\lambda I + \Phi^T\Phi)^{-1} = (\lambda I)^{-1}\Phi^T(I^{-1} + \Phi(\lambda I)^{-1}\Phi^T)^{-1}(I)^{-1}(\Phi^T)^{-1}$$

$$(\lambda I + \Phi^T\Phi)^{-1} = \frac{1}{\lambda}\Phi(I + \frac{1}{\lambda}\Phi\Phi^T)^{-1}(\Phi^T)^{-1}$$

$$(\lambda I + \Phi^T\Phi)^{-1}\Phi^T = \Phi^T(\lambda I + \Phi\Phi^T)^{-1}$$

$$\Rightarrow w = (\Phi^T\Phi + \lambda I)^{-1}\Phi^T t = \Phi^T(\lambda I + \Phi\Phi^T)^{-1}t = \Phi^T a,$$

where $a = (\lambda I + \Phi\Phi^T)^{-1}t$

$$f(x) = w^T\phi(x) = (\Phi^T a)^T\phi(x) = a^T\Phi(a)\phi(x) = a^T k(x)$$

where $k(x) = \Phi\phi(x) = [k(x_1, x), \dots k(x_n, x)]^T$

$$E(w) = (\Phi w - t)^T(\Phi w - t) + \lambda w^T w$$

$$= w^T\Phi^T\Phi w - 2t^T\Phi w + t^T t + \lambda a^T\Phi\Phi^T a$$

$$= a^T KKa - 2t^T Ka + t^T t + \lambda a^T Ka$$