Math 597 Assignment 2

Jonathan Pearc
260672004

1) (1.1) $y(x, w) = \sum_{j=0}^{M} w_j x^j$

(1.2) $E(w) = \frac{1}{2} \sum_{n=1}^{N} \left(y(x_n, w) - t_n \right)^2$

$$E(w) = \frac{1}{2} \sum_{n=1}^{N} \left(\sum_{j=0}^{M} w_j x_n^j - t_n \right)^2$$

$$\Rightarrow \quad \frac{\partial E(w)}{\partial w_i} = \sum_{n=1}^{N} \left[\sum_{j=0}^{M} w_j x_n^j - t_n \right] x_n^i = 0$$

$$\Rightarrow \quad \sum_{n=1}^{N} \left[\sum_{j=0}^{M} w_j x_n^j \right] x_n^i = \sum_{n=1}^{N} t_n x_n^i$$

$$\Rightarrow \quad \sum_{n=1}^{N} \sum_{j=0}^{M} w_j x_n^{i+j} = \sum_{n=1}^{N} t_n x_n^i$$

$$\Rightarrow \quad \sum_{j=0}^{M} w_j \sum_{n=1}^{N} x_n^{i+j} = \sum_{n=1}^{N} x_n^i t_n$$

let $A_{ij} = \sum_{n=1}^{N} x_n^{i+j}$, $T_i = \sum_{n=1}^{N} x_n^i t_n$

$$\Rightarrow \quad \boxed{\sum_{j=0}^{M} A_{ij} w_j = T_i}$$

2) (3.29) $\frac{1}{2}\sum_{n=1}^{N}(t_n - w^T\phi(x_n))^2 + \frac{\lambda}{2}\sum_{j=1}^{M}|w_j|^q$

(3.12) $\frac{1}{2}\sum_{n=1}^{N}(t_n - w^T\phi(x_n))^2$

(3.30) $\sum_{j=1}^{M}|w_j|^q \leq \eta$

we will first manipulate the constraint

(3.30) $\iff \sum_{j=1}^{M}|w_j|^q - \eta \leq 0$

$\Rightarrow \frac{1}{2}\left(\sum_{j=1}^{M}|w_j|^q - \eta\right) \leq 0$ Scaling dues not effect the constraint

Combine constraint with (3.12) to obtain Lagrange function

$$\boxed{L(w,\lambda) = \frac{1}{2}\sum_{n=1}^{N}(t_n - w^T\phi(x_n))^2 + \frac{\lambda}{2}\left(\sum_{j=0}^{M}|w_j|^q - \eta\right)}$$

this equation is equivilant to (3.29) with respect to w

If we were to minimize $L(w,\lambda)$ and denote the computed w value w^*, using the 3rd KKT condition, we can compute η

$$\lambda g(x) = 0$$

$$\Rightarrow \frac{\lambda}{2}\left(\sum_{j=0}^{M}|w_j^*|^q - \eta\right) = 0$$

$$\boxed{\Rightarrow \eta = \sum_{j=0}^{M}|w_j^*|^q}$$

3) let v be a n dimensional vector in $\mathbb{R}^n$

let p be the projection of v in the column space of Φ

$$\Rightarrow p = \phi_1(x) y_1 + \dots + \phi_M(x) y_M = \Phi y \qquad y \in \mathbb{R}^M$$

$v = p + e$ where e is in the orthogonal complement of the column space of Φ

$\Rightarrow$ e is in the null space of Φ^T

$\Rightarrow e = v - p$

$\Rightarrow$ $v - p$ is also in the null space of Φ^T

$$\Rightarrow \Phi^T (v - p) = 0 \qquad \Rightarrow \Phi^T v = \Phi^T \Phi y$$

$$\Rightarrow \Phi^T v - \Phi^T p = 0$$

$$\Rightarrow \Phi^T v - \Phi^T \Phi y = 0$$

Assuming columns of $\Phi^T \Phi$ are linearly indepedent, then $\Phi^T \Phi$ is invertible

$$\Rightarrow y = (\Phi^T \Phi)^{-1} \Phi v$$

$$\Rightarrow p = \Phi y = \Phi (\Phi^T \Phi)^{-1} \Phi^T v$$

$\boxed{\Rightarrow \Phi(\Phi^T \Phi)^{-1} \Phi^T \text{ projects } \sout{\text{and}} \text{ vector } v \text{ onto the column space of } \Phi}$

· further, suppose $t \in \mathbb{R}^n$

$$y = (\Phi^T \Phi)^{-1} \Phi^T t = \underbrace{w_{ML}}_{\text{from (3.15)}} t$$

$\Rightarrow$ y is the least squares solution

$\Rightarrow$ Least squares solution corresponds to an orthogonal projection of the vector t

4) $a_n = -\frac{1}{\lambda}(w^T \phi(x_n) - t_n)$

$$= \frac{-1}{\lambda}\left(w_1 \phi_1(x_n) + \dots + w_M \phi_M(x_n) - t_n\right)$$

$$= \frac{-w_1}{\lambda}\phi_1(x_n) - \dots - \frac{w_M}{\lambda}\phi_M(x_n) + \frac{t_n}{\lambda}$$

$$= \frac{-w_1}{\lambda}\phi_1(x_n) - \dots - \frac{w_M}{\lambda}\phi_M(x_n) + \frac{t_n}{\lambda} \cdot \frac{[\phi_1(x_n) + \dots + \phi_M(x_n)]}{\underbrace{[\phi_1(x_n) + \dots + \phi_M(x_n)]}_{\text{denote } S_\phi}}$$

$$= \frac{-w_1}{\lambda}\phi_1(x_n) - \dots - \frac{w_M}{\lambda}\phi_M(x_n) + \frac{t_n}{\lambda S_\phi}\phi_1(x_n) + \dots + \frac{t_n}{\lambda S_\phi}\phi_M(x_n)$$

$$= \left(\frac{t_n}{\lambda S_\phi} - \frac{w_1}{\lambda}\right)\phi_1(x_n) + \dots + \left(\frac{t_n}{\lambda S_\phi} - \frac{w_M}{\lambda}\right)\phi_M(x_n)$$

$$= \sum_{i=1}^{M}\left(\frac{t_n}{\lambda S_\phi} - \frac{w_i}{\lambda}\right)\phi_i(x_n)$$

$\Rightarrow$ a_n is a linear combination of the elements of the vector $\phi(x_n)$

let $v_i = \frac{t_n}{\lambda S_\phi} - \frac{w_i}{\lambda}$

$$\Rightarrow a_n = \sum_{i=1}^{M} v_i \phi_i(x_n)$$

$\boxed{\Rightarrow a = \Phi v}$ where $v = (v_1 \dots v_i \dots v_M)$

$$J(a) = \frac{1}{2} a^T K K a - a^T K t + \frac{1}{2} t^T t + \frac{\lambda}{2} a^T K a \qquad (6.7)$$

Sub in $a = \Phi v$

$$= \frac{1}{2} v^T \Phi^T K K \Phi v - v^T \Phi^T K t + \frac{1}{2} t^T t + \frac{\lambda}{2} v^T \Phi^T K \Phi v$$

$$= \frac{1}{2}\left[(K\Phi v)^T (K\Phi v) - 2 v^T \Phi^T K t + t^T t\right] + \frac{\lambda}{2} v^T \Phi^T K \Phi v$$

$$= \frac{1}{2}\left(K\Phi v - t\right)^T \left(K\Phi v - t\right) + \frac{\lambda}{2} v^T \Phi^T K \Phi v$$

$$= \frac{1}{2}\left(\Phi \underbrace{\Phi^T \Phi v}_{} - t\right)^T \left(\Phi \underbrace{\Phi^T \Phi v}_{} - t\right) + \frac{\lambda}{2} \underbrace{v^T \Phi^T \Phi}_{} \underbrace{\Phi^T \Phi v}_{}$$

recall $w = \Phi^T a = \Phi^T \Phi v \qquad (w^T = v^T \Phi^T \Phi)$

$$= \frac{1}{2}\left(\Phi w - t\right)^T \left(\Phi w - t\right) + \frac{\lambda}{2} w^T w$$

This is equivilant to 6.2

5) (6.15) : $k(x, x') = g(k_1(x, x'))$

where $g(x) = a_0 + a_1 x + a_2 x^2 + \dots a_n x^n = \sum_{i=0}^{n} a_i x^i$

$$k(x, x') = g(k_1(x, x'))$$

$$= \sum_{i=0}^{n} a_i [k_1(x, x')]^i$$

Using (6.13), (6.17), (6.18) this summation can be reduced ~~and~~ to a single kernel, therefore showing $k(x, x')$ is a valid kernel.

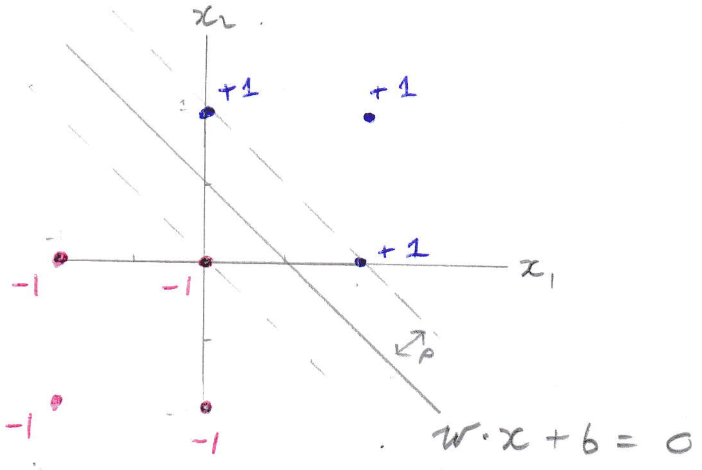
(6.16) : $k(x, x') = \exp(k_1(x, x'))$

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots = \sum_{i=0}^{\infty} \frac{x^i}{i!}$$

$$\exp(k_1(x, x')) = \sum_{i=0}^{\infty} \frac{k_1(x, x')^i}{i!}$$

If we take $g(k_1(x, x'))$ from (6.15) and let $a_i := \frac{1}{i!} \geq 0$ and then take $n \to \infty$ we get exactly $\exp(k(x, x_i))$. Therefore (6.13), (6.17), (6.18) can be used again to reduce $\exp(k_1(x, x$ to a single kernel. Therefore $k(x, x')$ is a valid kernel

6) a)



b) $x = (x_1, x_2)$
$w = (w_1, w_2)$

$$\min_{w,b} \|(w_1, w_2)\|^2 \quad \text{subject to} \quad \begin{cases} (w_1 + w_2 + b) \geq 1 \\ (w_1 + b) \geq 1 \\ (w_2 + b) \geq 1 \\ -(-w_1 - w_2 + b) \geq 1 \\ -(-w_1 + b) \geq 1 \\ -(w_2 + b) \geq 1 \\ -(b) \geq 1 \end{cases}$$

c) The line $w \cdot x + b = 0$ in part a) has form $0 = x_1 + x_2 - 0.5 \iff w = (1,1)$, $b = -0.5$. However note in part b) $-b \geq 1$. Scaling the equation by 2 we get $w = (2,2)$ $b = -1$, these satisfy all 7 point constraints, while making w minimum.

$$\|w\|_2^2 = 2^2 + 2^2 = 8 \implies \frac{1}{\|w\|_2^2} = \frac{1}{8}$$

$$\rho = \frac{\sqrt{(\frac{1}{2})^2 + (\frac{1}{2})^2}}{2} = \frac{\sqrt{\frac{1}{2}}}{2} = \frac{1}{2\sqrt{2}} \implies \rho^2 = \frac{1}{8}$$

$$\boxed{\implies \rho^2 = \frac{1}{\|w\|^2}}$$