PRML Test 1 (EE22MTECH02003)

Ans (2) Given, $\bar{\omega} = \begin{pmatrix} 1 \\ \epsilon \end{pmatrix} \in \mathbb{R}^2$

so,

$$\|w\|_1 = 1 + \epsilon \longrightarrow L_1 \text{ Norm}$$

$$\|w\|_2 = 1 + \epsilon^2 \longrightarrow L_2 \text{ Norm}$$

Now, comparing the results obtained by reducing $\bar{\omega}$ by a small value $\begin{pmatrix} \alpha \\ 0 \end{pmatrix}$ where $\alpha < \epsilon$

$$\left.\begin{aligned} \left\|\begin{pmatrix} 1 \\ \epsilon \end{pmatrix} - \begin{pmatrix} \alpha \\ 0 \end{pmatrix}\right\|_1 &= 1 - \alpha + \epsilon \\ \left\|\begin{pmatrix} 1 \\ \epsilon \end{pmatrix} - \begin{pmatrix} 0 \\ \alpha \end{pmatrix}\right\|_1 &= 1 - \alpha + \epsilon \end{aligned}\right\} L_1 \text{ Norms}$$

$$\left.\begin{aligned} \left\|\begin{pmatrix} 1 \\ \epsilon \end{pmatrix} - \begin{pmatrix} \alpha \\ 0 \end{pmatrix}\right\|_2 &= \sqrt{1 - 2\alpha + \alpha^2 + \epsilon^2} \\ \left\|\begin{pmatrix} 1 \\ \epsilon \end{pmatrix} - \begin{pmatrix} 0 \\ \alpha \end{pmatrix}\right\|_2 &= \sqrt{1 - 2\epsilon\alpha + \alpha^2 + \epsilon^2} \end{aligned}\right\} L_2 \text{ Norms}$$

We can see,

for L_2 norm, regularizing the larger term results in larger reduction in norm output when compared to smaller one.

Whereas for L_1 norm, reduction is independent of quantity being penalized and is always equal to α.

Hence, L_2 norm discourages sparsity.

Ans (1) Given, training set is $(\bar{x}_{1:N}, t_{1:N})$

using Least Squares approach, estimated model parameter is :-

$$\bar{\omega}(N) = \left(\phi^T(\bar{x})\,\phi(\bar{x})\right)^{-1} \phi(\bar{x}) t$$

$\Rightarrow$

Now, estimated target values for given additional M input points $(x_{N+1:N+M})$ can be regressed by:—

$$\hat{t} = y(\bar{x}, \bar{w}) = \bar{w}^{(N)T} \phi(\bar{x}_{N+1:N+M})$$

So,
modified train set is :- $(\bar{x}_{1:N+M}, \bar{t}_{1:N+M})$
and final error with new train set :—

$$J(\bar{w}) = \frac{1}{2} \sum_{n=1}^{M} (t_n - w^T \phi(x_n))^2$$

$$J(\bar{w}) = \frac{1}{2} \sum_{n=1}^{N} (t_n - w^T \phi(x))^2 + \frac{1}{2} \sum_{n=N+1}^{N+M} (\hat{t}_n - w^T \phi(x_n))^2$$

$$\nabla J(\bar{w}) = \underbrace{-\phi^T(x)(t - w^T \phi(x))}_{x = 1 \text{ to } N} - \underbrace{\phi^T(x)(w^{(N)T} \phi(x) - w^T \phi(x))}_{x = N+1 \text{ to } N+M}$$

$$= 0$$

$$\Rightarrow \phi^T(x)\phi(x) w^{N+M} = \phi^T(x) t + \phi^T(x)\phi(x) w^N$$

$$w^{N+M} = (\phi^T(x)\phi(x))^{-1} \phi^T(x) t + (\phi^T(x)\phi(x))^{-1} \times \phi^T(x)\phi(x) w^N$$

$$w^{N+M} = w^N + (\phi^T(x)\phi(x))^{-1} \phi^T(x) t$$

Hence, w^{N+M} performs better than w^N.

Ans 4 $\bar{x} = [0 \quad 1]^T$

$$\bar{t} = \left[\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ -2 \end{pmatrix}, \begin{pmatrix} -2 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right]$$

$$\phi(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \phi(1) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\left\{ w = (\phi^T(\bar{x})\phi(\bar{x}))^{-1} \phi^T(x) t \right\}$$

$$\phi(\bar{x}) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I_{2\times2} \qquad \phi^T(x)\phi(x) = I$$

Hence, $\boxed{\bar{w} = \bar{t}}$

Now, given $p(\bar{t}/\bar{x}, w) = \prod_{k=1}^{K} \mathcal{N}(t_k / w_k^T \phi_k(x) \beta_k^{-1})$

$$\log p(\bar{t}/\bar{x}, w) = -\frac{\beta}{2} \sum_{k=1}^{K} (t_k - w_k^T \phi_k(x))^2 - \frac{K}{2} \log \beta - \frac{k}{2} \log 2\pi$$

$$\max \log(p(\bar{t}/\bar{x}, w)) = \min \frac{1}{2} \sum_{k=1}^{K} (t_k - w_k^T \phi_k(x))^2$$

$$\Rightarrow w_{ML}^{(k)} = (\phi_k^T(x) \phi_k(x))^{-1} \phi_k(x) t_k$$

$$\beta_{ML}^{(k)} = \frac{1}{K} \sum_{k=1}^{K} (t_k - w_{ML}^T \phi_k(x))^2$$

$$w_{ML}(0) = \left[(1 \; 0) \begin{pmatrix} 1 \\ 0 \end{pmatrix}\right]^{-1} \begin{pmatrix} 1 \\ 0 \end{pmatrix} (-1 \; -1 \; -2 \; 1 \; 1 \; 2)$$

$$\therefore w_{ML}(0) = \begin{bmatrix} -1 & -1 & -2 & 1 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

& $w_{ML}(1) = \left[(0 \; 1) \begin{pmatrix} 0 \\ 1 \end{pmatrix}\right]^{-1} \begin{pmatrix} 0 \\ 1 \end{pmatrix} (-1 \; -2 \; -1 \; 1 \; 2 \; 1)$

$$w_{ML}(1) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -2 & -1 & 1 & 2 & 1 \end{bmatrix}$$

Ans (3) A/C, errors follow laplacian distribution

$\therefore p(t_n / \bar{x}_n, \bar{w}, \beta) = e_n \sim \text{Laplace}(0, b)$

$$= \frac{1}{2b} \exp\left(-\frac{e_n}{b}\right)$$

$\therefore$ mean $= b$, variance $= b^2$

(a) $J_L(\bar{w}) = \max -\frac{\beta}{2} \sum_{n=1}^{N} (t_n - \bar{w}^T \phi(x_n))^2$

(b) $\cancel{w_{MAP} = (\phi^T \phi)^{-1} \phi^T t}$ $\quad w_{MAP} = (\phi^T \phi + \lambda I)^{-1} \phi^T t$

where $\lambda = \frac{\alpha}{\beta}$ corresponding to L_1 regularizer regularization term is $\frac{\alpha}{\beta} |w_1| = \lambda |w_1|$