B 555 Machine Learning
Written Assignment 2

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

(i) To minimise this error fn.

$$\frac{d}{dw} E_D(w) = 0$$

noise variance $(\sqrt{B})$

$$\frac{1}{\cancel{2}} \times -\cancel{2} \sum_{n=1}^{N} r_n \{t_n - w^T\phi(x_n)\}\phi(x_n) = 0$$

$$\sum_{n=1}^{N} -r_n t_n \phi(x_n) + \sum_{n=1}^{N} r_n w^T\phi(x_n)\phi(x_n) = 0$$

$$\sum_{n=1}^{N} r_n \phi(x_n)^T w\, \phi(x_n) = \sum_{n=1}^{N} r_n t_n \phi(x_n)$$

$$\left\{\sum_{n=1}^{N} r_n \phi(x_n)\phi(x_n)^T\right\} w = \sum_{n=1}^{N} r_n t_n \phi(x_n)$$

$$\boxed{w = \left(\sum_{n=1}^{N} r_n \phi(x_n)\phi(x_n)^T\right)^{-1}\left(\sum_{n=1}^{N} r_n t_n \phi(x_n)\right)}$$

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

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

Replicated Datapts

Let $t_n = t$ & $\phi(x_n)^T = \Phi$

$= \cancel{\frac{1}{2} \sum_{n=1}^{N} r_1}$ Using matrix notation for error fn.

$= \frac{1}{2} R (t - \Phi w)^2$ [Here R is diagonal matrix of $(r_1, r_2 \ldots r_N)$)

$= \frac{1}{2} (t - \Phi w)^T R (t - \Phi w)$

$E_D(w) = \frac{1}{2} \left[t^T t - 2t^T R \Phi w + w^T \Phi^T R \Phi w \right]$

Taking gradient of $E_D(w)$

$\nabla E_D(w) = \frac{1}{\cancel{2}} \times -\cancel{2} \times t^T R \Phi + \frac{1}{\cancel{2}} \times \cancel{2} \Phi^T R \Phi w = 0$

$\Rightarrow (\Phi^T R \Phi) w = t^T R \Phi$

$\Rightarrow w = (\Phi^T R \Phi)^{-1} t^T R \Phi$

$\Rightarrow \boxed{w = (\Phi^T R \Phi)^{-1} \cdot \Phi^T R t}$

Ans.2) $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim \text{Uniform } [1, 2]^2$

$y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_2^2 \\ x_1 + 5x_2 \end{pmatrix}$ & $y_1 \in [1, 4]$

& $y_2 \in [6, 12]$

$$J = \begin{vmatrix} \frac{d(x_2)}{y_1} & \frac{d(x_2)}{y_2} \\ \frac{d(x_1)}{y_1} & \frac{d(x_1)}{y_2} \end{vmatrix}$$

$x_2^2 = y_1$ & $x_1 + 5x_2 = y_2$

$x_2 = \sqrt{y_1}$ & $x_1 + 5\sqrt{y_1} = y_2$

& $x_1 = y_2 - 5\sqrt{y_1}$

$$J = \begin{vmatrix} \frac{1}{2} y_1^{-\frac{1}{2}} & 0 \\ -\frac{5}{2} y_1^{-\frac{1}{2}} & 1 \end{vmatrix}$$

$$\boxed{|J| = \frac{1}{2} y_1^{-\frac{1}{2}}}$$

Now $f(y) = f(x) \times |J|$

$$f(x) = f(x_1) \cdot f(x_2) = \begin{cases} \frac{1}{2-1} & 1 < x_1, x_2 < 2 \\ 0 & o \cdot w \end{cases}$$

$$f(y) = \begin{cases} \frac{1}{2} y_1^{-\frac{1}{2}} & 1 < y_1 < 4 \\ 0 & o \cdot w \end{cases}$$

(iii) $\int_{-\infty}^{\infty} f(y) = \int \frac{1}{2} y_1^{-\frac{1}{2}} dy$

$$= \left[y^{\frac{1}{2}} \right]_1^4$$

$$= 4^{\frac{1}{2}} - 1^{\frac{1}{2}}$$

$$= \boxed{1}$$

Ans. 5. Verify: $P(w|t) = N(w \mid m_N, S_N)$ → Posterior

where $m_N = S_N (S_0^{-1} m_0 + \beta \Phi^T t)$

$$S_N^{-1} = S_0^{-1} + \beta \Phi^T \Phi$$

- Prior: $P(w) = N(w \mid m_0, S_0) = \sqrt{\frac{S_0^{-1}}{2\pi}} \times \exp\left\{ \frac{-S_0^{-1}}{2} \left[w - m_0 \right]^2 \right\}$

Likelihood : $P(t|\overset{X}{\wedge}w, \beta) = \prod_{n=1}^{N} \mathcal{N}\left(t_n | w^T \phi(x_n), \beta^{-1}\right)$

$$L = \sqrt{\frac{\beta}{2\pi}} \times \exp\left\{\frac{-\beta}{2} \sum_{n=1}^{N} \left(t_n - w^T \phi(x_n)\right)^2\right\}$$

let t_n be t

Also, we know $\phi(x_n)^T = \Phi$ (Design Matrix)

$$L = \sqrt{\frac{\beta}{2\pi}} \times \exp\left\{\frac{-\beta}{2}\left(t - \Phi w\right)^2\right\}$$

Posterior $\propto$ Prior $\times$ Likelihood

Considering only the exponent terms & ignoring the constants

$$\text{Posterior} \propto \exp\left\{-\frac{S_0^{-1}}{2}(w-m_0)^T(w-m_0)\right\} \times \exp\left\{\frac{-\beta}{2}(t-\Phi w)^T(t-\Phi w)\right\}$$

$$\propto \exp\left\{\frac{-S_0^{-1}}{2}\left(w^T w - 2m_0^T w + m_0^T m_0\right)\right\} \times \exp\left\{\frac{-\beta}{2}\left(t^T t - 2t^T \Phi w + w^T \Phi^T \Phi w\right)\right\}$$

$$\propto \exp\left\{\frac{-1}{2}\left[S_0^{-1} w^T w - 2S_0^{-1} m_0^T w + S_0^{-1} m_0^T m_0 + \beta t^T t - 2\beta t^T \Phi w + \beta w^T \Phi^T \Phi w\right]\right\}$$

$$\propto \exp\left\{\frac{-1}{2} w^T\left[S_0^{-1} + \beta \Phi^T \Phi\right] w - \frac{1}{2}\left[-2m_0^T S_0^{-1} - 2\beta t \Phi^T\right] w + \text{Constant terms}\right\}$$

$\propto \exp \{$ Comparing with the standard values for S_N & M_n for standard Gaussian

We get,

Compare quadratic term

$$\boxed{S_N^{-1} = S_0^{-1} + \beta \Phi^T \Phi}$$

Compare Linear Term

$$-2 m_N^T S_N^{-1} = -2 m_0^T S_0^{-1} - 2 \beta t \Phi^T$$

Cancelling out -2's.

$$m_N^T S_N^{-1} = m_0^T S_0^{-1} + \beta \Phi^T t$$

Taking Transpose both sides

$$\boxed{m_N = \frac{1}{S_N^{-1}} \left(S_0^{-1} m_0 + \beta \Phi^T t \right)} \rightarrow t^T \Phi = \Phi^T t$$

Ans 6 The uncertainity is given for predictive distn as:

$$\sigma_N^2(x) = \underbrace{\frac{1}{\beta}}_{\text{noise on data}} + \underbrace{\phi(x)^T S_N \phi(x)}_{\text{uncertainity associated with parameters } w} - (A)$$

To prove: $\sigma_{N+1}^2(x) \leq \sigma_N^2(x)$

Using:

$$(M + vv^T)^{-1} = M^{-1} - \frac{(M^{-1}v)(v^T M^{-1})}{1 + v^T M^{-1} v} - (B)$$

We know,

$$S_N^{-1} = S_0^{-1} + \beta \Phi^T \Phi$$

$$= S_0^{-1} + \beta \sum_{n=1}^{N} \phi(x_n)\,\phi(x_n)^T \quad -① $$

Now,

$$S_{N+1}^{-1} = S_0^{-1} + \beta \sum_{n=1}^{N+1} \phi(x_n)\,\phi(x_n)^T$$

Using ①

$$S_{N+1}^{-1} = S_N^{-1} + \beta\,\phi(x_{N+1})(x_{N+1})^T \quad -③$$

Now,

$$\sigma_{N+1}^2(x) = \frac{1}{\beta} + \phi(x)^T S_{N+1}\,\phi(x)$$

Using Ⓐ

$$\Rightarrow \sigma_{N+1}^2(x) = \sigma_N^2(x) - \phi(x)^T S_N\,\phi(x) + \phi(x)^T S_{N+1}\phi(x)$$

$$\Rightarrow \sigma_{N+1}^2(x) = \sigma_N^2(x) + \phi(x)^T\left(S_{N+1} - S_N\right)\phi(x)$$

$$\Rightarrow \frac{\sigma_{N+1}^2(x) - \sigma_N^2(x)}{\phi(x)^T\,\phi(x)} = S_{N+1} - S_N \quad — ②$$

Putting $M = S_N^{-1}$ & $V = \sqrt{\beta}\,\phi(x_{N+1})$ in L.H.S of Ⓑ

$$= \left(S_N^{-1} + \beta\,\phi(x_{N+1})\,\phi(x_{N+1})^T\right)^{-1}$$

Using ③

$$= \left(S_{N+1}^{-1}\right)^{-1}$$

$$= S_{N+1} \quad - \text{L.H.S}$$

Substituting values of M & V in R.H.S of (B)

R.H.S

$$= S_N - \frac{(S_N).(\sqrt{\beta}\,\phi(x_{N+1}))\,(\sqrt{\beta}\,\phi(x_{N+1})^T)\,S_N}{1+\sqrt{\beta}\,\phi(x_{N+1})^T.\,S_N.\,\sqrt{\beta}\,\phi(x_{N+1})}$$

$$S_{N+1} = S_N - \frac{\beta S_N \phi(x_{N+1}).\phi(x_{N+1})^T S_N}{1+\beta\,\phi(x_{N+1})^T\phi(x_{N+1})}$$

Putting L.H.S & R.H.S together

$$S_{N+1} - S_N = \frac{-\beta S_N \phi(x_{N+1}).\phi(x_{N+1})^T S_N}{1+\beta\,\phi(x_{N+1})^T\phi(x_{N+1})}$$

From (2)

$$\frac{\sigma^2_{N+1}(x) - \sigma^2_N(x)}{\phi(x)^T\phi(x)} = \frac{-\beta S_N \phi(x_{N+1}).\phi(x_{N+1})^T S_N}{1+\beta\,\phi(x_{N+1})^T\phi(x_{N+1})}$$

$$\sigma^2_{N+1}(x) = \sigma^2_N(x) - \frac{\beta\,\phi(x)^T S_N \phi(x_{N+1}).\phi(x_{N+1})^T S_N.\phi(x)}{1+\beta\phi(x_{N+1})^T S_N \phi(x_{N+1})}$$

Now S_N is positive definite

Also, $S_N\,\phi(x_{N+1})\,\phi(x_{N+1})^T$ is positive semi-definite, the 2nd term becomes non negative & Hence

$$\boxed{\sigma^2_{N+1}(x) \le \sigma^2_N(x)}$$

Ans 3. $x \sim N(0, I)$

Comparing with:

$x \sim N(\mu, \Lambda^{-1})$

$\therefore \mu = 0$

$\Lambda^{-1} = I$

$\therefore \Lambda = I$

$y|x \sim N(\mu = 3x_1 + 2x_2 + 5, \sigma^2 = 25)$

Comparing with:

$y|x \sim N(Ax + b, L^{-1})$

So, $A = \begin{bmatrix} 3 & 2 \end{bmatrix}$; $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$b = 5$; $L^{-1} = 25$

$\therefore L = \frac{1}{25}$

$P(x|y=4) = N(\mu^*, \Sigma^*)$

Comparing with:

$$\Sigma^* = (\Lambda + A^T L A)^{-1}$$

$$= \left(I + \begin{bmatrix} 3 \\ 2 \end{bmatrix} \cdot \frac{1}{25} \cdot \begin{bmatrix} 3 & 2 \end{bmatrix} \right)^{-1}$$

$$= \left(I + \begin{bmatrix} 3 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 3/25 & 2/25 \end{bmatrix} \right)^{-1}$$

$$= \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 9/25 & 6/25 \\ 6/25 & 4/25 \end{bmatrix} \right)^{-1}$$

$$\Sigma^* = \begin{pmatrix} 34/25 & 6/25 \\ 6/25 & 29/25 \end{pmatrix}^{-1}$$

Now $|\Sigma^*|^{-1} = \frac{38}{25}$

$$\Sigma^* = \begin{pmatrix} 29/38 & -6/38 \\ -6/38 & 34/38 \end{pmatrix}$$

x_1 & x_2 are not independent anymore $\because$ the covariance matrix is not a diagonal matrix anymore which is clearly visible from Σ^*

$$\mu^* = \Sigma^* \left[A^T L (y-b) + \Lambda \mu \right]$$

$$= \begin{bmatrix} 29/38 & -6/38 \\ -6/38 & 34/38 \end{bmatrix} \left(\begin{bmatrix} 3 \\ 2 \end{bmatrix} \cdot \frac{1}{25} (4-5) + I \times 0 \right)$$

$$= \begin{bmatrix} 29/38 & -6/38 \\ -6/38 & 34/38 \end{bmatrix} \left(\begin{bmatrix} 3 \\ 2 \end{bmatrix} \cdot \left(\frac{-1}{25} \right) \right)$$

$$= \begin{bmatrix} 29/38 & -6/38 \\ -6/38 & 34/38 \end{bmatrix}_{2\times 2} \begin{bmatrix} -3/25 \\ -2/25 \end{bmatrix}_{2\times 1}$$

$$= \begin{bmatrix} \frac{-87+12}{38\times 25} \\ \\ \frac{18-68}{38\times 25} \end{bmatrix} = \begin{bmatrix} \frac{-\cancel{75}^{\,3}}{38\times \cancel{25}_{1}} \\ \frac{-\cancel{50}^{\,2}\,{}^{1}}{{}_{19}\cancel{38}\times \cancel{25}_{1}} \end{bmatrix}$$

$$\mu^* = \begin{bmatrix} -3/38 \\ -1/19 \end{bmatrix}$$

Ans 4 S is a symmetric matrix with eigen decomposition $S = V \Lambda V^T$

Now

$$\arg\max_{\{x \mid x^T x \le 1\}} x^T S x$$

$$L(x, \lambda) = f(x) + \lambda(g(x))$$

$$L = x^T S x + \lambda(1 - x^T x)$$

Differentiate w.r.t x

$$\frac{dL}{dx} = 2Sx - 2\lambda x$$

Putting it to 0

$$2Sx = 2\lambda x$$

$$\boxed{Sx = \lambda x}$$

$\therefore$ (λ, x) is the eigenvalue-eigenvector pair of S which corresponds to (Λ_i, v_i)

Now $\arg\max_{x \mid x^T x \le 1} x^T S x$

$$= \arg\max_{x \mid x^T x \le 1} x^T \lambda x = \arg\max_{x \mid x^T x \le 1} \lambda x^T x$$

Now it's given that $x^T x \le 1$ $\therefore$ $\max(x^T x) = 1$

& hence $\arg\max_{x \mid x^T x \le 1} \lambda x^T x = \arg\max \lambda$

$$\arg\max_{x \mid x^T x \le 1} \lambda x^T x = \Lambda_{max} I$$

Ans.7. Pot $\frac{d}{dA} \ln|A| = (A^{-1})^T$

$$= \frac{1}{|A|} \times \frac{d}{dA_{ij}} |A|$$

Using the Jacobian

$$= \frac{1}{|A|} \times \frac{d}{dA_{ij}} \sum_i \sum_j A_{ij} \times C_{ij}(A)$$

$$= \frac{1}{|A|} \times \sum_i \sum_j C_{ij}(A)$$

$$= \frac{1}{|A|} \times Adj^T(A)$$

We know, $A^{-1} = \frac{Adj(A)}{|A|}$

Transposing both sides

$$(A^{-1})^T = \frac{Adj^T(A)}{|A|} \quad \text{— (1)}$$

$= \boxed{(A^{-1})^T}$ Using (1)

Hence proved