HOMEWORK 2

Question 1.1:

$X = (X_1 \dots X_n)$, X_i = delay in minutes

Sol^n:

i) To find log likelihood function of X given λ

$$\therefore \hat{\lambda} = \underset{\lambda}{\arg\max} \prod_{k_i=1}^{n} P(x_i = k \mid \lambda)$$

$$= \prod_{k_i=1}^{n} \frac{\lambda^{k_i}}{k_i!} e^{-\lambda}$$

$\Rightarrow$ Taking log (natural) on both sides we get:

$$\text{Log likelihood} = \ln \left[\prod_{k_i=1}^{n} \frac{\lambda^{k_i}}{k_i!} e^{-\lambda} \right]$$

$$= \ln \left[\frac{\lambda^{k_1}}{k_1!} \cdot \frac{\lambda^{k_2}}{k_2!} \cdots \frac{\lambda^{k_n}}{k_n!} e^{-n\lambda} \right]$$

$$= \ln e^{-n\lambda} + \ln \frac{\lambda^{k_1}}{k_1!} + \ln \frac{\lambda^{k_2}}{k_2!} \cdots \ln \frac{\lambda^{k_n}}{k_n!}$$

$$\therefore \text{Log likelihood} = -n\lambda + [k_1 \ln\lambda + k_2 \ln\lambda \cdots k_n \ln\lambda - \ln k_1 k_2 \cdots k_n]$$

$$= \boxed{-n\lambda + [(k_1+k_2+k_3 \cdots k_n)\ln\lambda - \ln k_1 k_2 \cdots k_n]}$$

ii) For MLE:

We differentiate the above eq^n and set it to 0.

$$\text{MLE} \Rightarrow \frac{\partial}{\partial\lambda} [-n\lambda + (k_1+k_2 \cdots k_n)\ln\lambda - \ln k_1 k_2 \cdots k_n] = 0$$

$$\Rightarrow -n + \frac{k_1+k_2 \cdots k_n}{\lambda} = 0$$

$$\therefore \boxed{\hat{\lambda} = \frac{\sum_{i=1}^{n} k_i}{n}}$$

Scanned by CamScanner

iii) MLE for λ using observed x:

$$\hat{\lambda} = \frac{1}{7}\left[4+5+3+5+6+9+10\right]$$

$$\therefore \hat{\lambda} = \frac{1}{7} \times 42 = \underline{\underline{6}}$$

Question 1.2:

i) For Posterior distribution, we know that:

$$P(\lambda|X) \propto P(X|\lambda) \cdot P(\lambda)$$

$$\propto \left(\prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda}\right) \cdot \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

$$\propto \frac{1}{\prod_{i=1}^{n} x_i!} \lambda^{\sum_{i=1}^{n} x_i} \cdot e^{-n\lambda} \cdot \frac{\beta^{\alpha}}{\Gamma(\alpha)} \lambda^{\alpha-1} e^{-\beta\lambda}$$

$$\propto \left(\frac{\beta^{\alpha}}{\Gamma(\alpha) \prod_{i=1}^{n} x_i!}\right) \lambda^{\left[\sum_{i=1}^{n} x_i + \alpha - 1\right]} e^{-\beta\lambda - n\lambda}$$

$$\propto [\text{const.}]\, \lambda^{[n \cdot x_{mean} + \alpha - 1]} e^{-\lambda[n+\beta]}$$

$$\propto \text{Gamma}\left[n x_{\mu} + \alpha - 1, n+\beta\right]$$

Taking log we get:

$$\log P(\lambda|X) = \log C + \log \lambda^{[n \cdot x_{\mu} + \alpha - 1]} + \log e^{-\lambda[n+\beta]}$$

$$= \log C + [n \cdot x_{\mu} + \alpha - 1] \log \lambda - \lambda [n+\beta]$$

ii) For MAP:

we differentiate above eq^n 8 set it to 0.

$$\therefore \text{MAP} \Rightarrow \frac{\partial}{\partial \lambda}\left[\log C + (n \cdot x_{\mu} + \alpha - 1) \log \lambda - \lambda[n+\beta]\right] = 0$$

$$= \frac{n \cdot x_{\mu} + \alpha - 1}{\lambda} - [n+\beta] = 0$$

$$\therefore \text{MAP} \Rightarrow \boxed{\hat{\lambda}_{MAP} = \frac{n \cdot x_{mean} + \alpha - 1}{n + \beta}}$$

Question 1.3:

i) Let $h = e^{-2\lambda}$, To show that $\hat{h} = e^{-2X}$ is MLE of h

$\therefore P(X|\lambda) = \dfrac{\lambda^X}{X!} e^{-\lambda}$

$h = e^{-2\lambda} \quad \ln h = -2\lambda$

$\therefore \lambda = -1/2 \ln h$

$\therefore \lambda = \ln 1/\sqrt{h}$

$P(X|h) = \dfrac{(-1/2 \ln h)^X}{X!} e^{-(-1/2 \ln h)}$ [ln = natural log]

$= \dfrac{(-1/2 \ln h)^X}{X!} h^{1/2}$

$\therefore \ln P(X|h) = \ln\left[\dfrac{(-1/2 \ln h)^X}{X!} h^{1/2}\right]$

$= \dfrac{1}{2}\ln h - \ln X! + \ln\left[(-1/2 \ln h)^X\right]$

$= \dfrac{1}{2}\ln h - \ln X! + X \ln(\ln 1/\sqrt{h})$

Differentiating & equating to 0

$\therefore$ MLE $\Rightarrow \dfrac{\partial}{\partial h}\left[\dfrac{1}{2}\ln h - \ln X! + X \ln \ln 1/\sqrt{h}\right] = 0$

$\Rightarrow \dfrac{1}{2h} - 0 + X\left[\dfrac{1}{\ln h^{-1/2}} \cdot \dfrac{1}{h^{-1/2}} \cdot -\dfrac{1}{2} h^{-3/2}\right] = 0$

$\dfrac{1}{2h} = \dfrac{X}{2}\left[\dfrac{1}{\ln h^{-1/2}} \cdot h^{1/2} \cdot \dfrac{1}{h^{3/2}}\right]$

$\therefore \dfrac{1}{\cancel{2}\cancel{h}} = \dfrac{X}{\cancel{2}} \cdot \dfrac{1}{\ln h^{-1/2}} \cdot \dfrac{1}{\cancel{h}}$

$\therefore \ln h^{-1/2} = X \Rightarrow \ln h = -2X$

$\therefore \boxed{\hat{h} = e^{-2X}}$

Hence proved.

Scanned by CamScanner

ii) Bias = Estimated value − True value

$\therefore$ $\text{bias}(\hat{\theta}) = E(\hat{\theta}) - \theta$

$\text{bias}(\hat{h}) = E(\hat{h}) - h$

$= E[e^{-2X}] - e^{-2\lambda}$

$\therefore$ $\text{Bias} = \left(\sum_x e^{-2X} \cdot \frac{\lambda^X e^{-\lambda}}{X!}\right) - e^{-2\lambda}$ [using LOTUS]

~~[E(g(h))]~~ =

$\therefore$ $\text{Bias} = e^{-\lambda}\left[\sum_x \frac{(\lambda/e^2)^X}{X!}\right] - e^{-2\lambda}$

we know that: $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$

$\therefore$ $\text{Bias} = e^{-\lambda}\left[e^{\lambda/e^2}\right] - e^{-2\lambda}$

$= e^{\lambda/e^2 - \lambda} - e^{-2\lambda}$

$\therefore$ $\boxed{\text{Bias} = e^{-\lambda[1 - 1/e^2]} - e^{-2\lambda}}$

iii) For unbiased estimate, let our f^n be ~~e^{mx}~~ m^X

$\therefore$ For unbiased estimate:

Expected value = true value

$\cancel{E(e^{mx}) = e^{-2\lambda}}$ $E(m^X) = e^{-2\lambda}$

$\cancel{\sum_x e^{mX} \frac{\lambda^X}{X!} e^{-\lambda} = e^{-2\lambda}}$ $\sum_x m^X \frac{\lambda^X}{X!} e^{-\lambda} = e^{-2\lambda}$

$\cancel{e^{-\lambda}\left[\sum_x (\right.}$ $\therefore$ $e^{-\lambda}\left[\sum_x \frac{(m\lambda)^X}{X!}\right] = e^{-2\lambda}$

$e^{m\lambda} = e^{-\lambda}$

$\therefore$ $m\lambda = -\lambda$

$m = -1$

$\therefore$ $(-1)^X$ is ~~so~~ the only unbiased estimate of h.

Question 2:

2.1) $\underset{w,b}{\text{minimize}} \; \lambda \|w\|^2 + \sum_{i=1}^{n} (w^T x_i + b - y_i)^2$

Let $\bar{w} = [w; b]$ & $\bar{X} = [X; 1_n^T]$

$\therefore$ we can write above f^n as:

$\underset{\bar{w}}{\min} \; \lambda [\bar{w}^T \cdot \bar{w} - b^2] + \sum_{i=1}^{n} (\bar{w}^T x_i - y_i)^2$

we differentiate the above eq^n & set it to 0.

$\frac{\partial}{\partial \bar{w}} [\lambda (\bar{w}^T \cdot \bar{w} - b^2] + \sum_{i=1}^{n} (x_i^T \bar{w} - y_i^T)^2] = 0$

[transpose of scalar is equal to the original & $(\bar{w}^T x_i - y_i)$ is a scalar]

$\therefore \lambda [2 \bar{I} \bar{w}] + \frac{\partial}{\partial \bar{w}} [\|\bar{X}^T \cdot \bar{w} - y\|^2] = 0$

$\therefore 2\lambda \bar{I} \bar{w} + \frac{\partial}{\partial \bar{w}} [(\bar{X}^T \bar{w} - y)^T (\bar{X}^T \bar{w} - y)] = 0$

$\therefore 2\lambda \bar{I} \bar{w} + \frac{\partial}{\partial \bar{w}} [(\bar{w}^T \bar{X} - y^T)(\bar{X}^T \bar{w} - y)] = 0$

$2\lambda \bar{I} \bar{w} + \frac{\partial}{\partial \bar{w}} [\bar{w}^T \bar{X} \bar{X}^T \bar{w} - \underbrace{\bar{w}^T \bar{X} y}_{\text{[scalar, hence take transpose]}} - y^T \bar{X}^T \bar{w} + y^T y] = 0$

$2\lambda \bar{I} \bar{w} + \frac{\partial}{\partial \bar{w}} [\bar{w}^T \bar{X} \bar{X}^T \bar{w} - 2 y^T \bar{X}^T \bar{w} + y^T y] = 0$

$2\lambda \bar{I} \bar{w} + [(\bar{X} \bar{X}^T + (\bar{X} \bar{X}^T)^T] \cdot \bar{w} - 2 \bar{X} y = 0$

$\therefore \cancel{2} \lambda \bar{I} \bar{w} + \cancel{2} \bar{X} \bar{X}^T \bar{w} = \cancel{2} \bar{X} y$

$[\bar{X} \bar{X}^T + \lambda \bar{I}] \bar{w} = \bar{X} y$

$\therefore \bar{w} = \underbrace{[\bar{X} \bar{X}^T + \lambda \bar{I}]^{-1}}_{C} \cdot \underbrace{\bar{X} y}_{d}$

$\therefore \boxed{\bar{w} = C^{-1} \cdot d}$

Hence proved

Scanned by CamScanner

ii) $C = \underset{(k+1)\times n}{\bar{X}}\ \underset{n\times(k+1)}{\bar{X}^T} + \lambda \underset{(k+1)\times(k+1)}{\bar{I}}$

$\therefore$ Initial dimensions of $C \Rightarrow (k+1)\times(k+1)$

even when we remove one column (say x_i)

$C = \bar{X}_i \bar{X}_i^T + \lambda \bar{I}_i \rightarrow (k+1)\times(k+1)$

$= (k+1)\times(n-1) \quad (n-1)\times(k+1)$

$\Rightarrow (k+1)\times(k+1)$

$\therefore$ dimensions remain same

when we see the matrix multiplications:

$$\begin{bmatrix} x_{1_1} & \cdots & x_{i_1} & \cdots & x_{n_1} \\ x_{1_2} & & x_{i_2} & & x_{n_2} \\ \vdots & & \vdots & & \vdots \\ x_{1_{k+1}} & & x_{i_{k+1}} & & x_{n_{k+1}} \end{bmatrix} \cdot \begin{bmatrix} x_{1_1} & \cdots & x_{1_{k+1}} \\ \vdots & & \vdots \\ x_{i_1} & & x_{i_{k+1}} \\ \vdots & & \vdots \\ x_{n_1} & & x_{n_{k+1}} \end{bmatrix}$$

in terms of column vectors:

$$[x_1\ x_2 \cdots x_i \cdots x_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{bmatrix}$$

$\therefore$ When we remove x_i, only terms missing from multiplication is $x_i \cdot x_i^T$

$\therefore$ $\boxed{C_{(i)} = C - \bar{x}_i \bar{x}_i^T}$

similarly for d:

$d = \bar{X} y$

$= [x_1\ x_2 \cdots x_i \cdots x_n] \cdot \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix}$

when we remove x_i, only term missing is $x_i \cdot y_i$

$\therefore$ $\boxed{d_{(i)} = d - \bar{x}_i y_i}$

iii) $C = (\bar{X}\bar{X}^T + \lambda\bar{I})$

$C^{-1} = (\bar{X}\bar{X}^T + \lambda\bar{I})^{-1}$

$\therefore C_{(i)} = (\bar{X}\bar{X}^T - \bar{x}_i\bar{x}_i^T + \lambda\bar{I})$

$\therefore C_{(i)} = [C - \bar{x}_i\bar{x}_i^T]$

$\therefore C_{(i)}^{-1} = [C - \bar{x}_i\bar{x}_i^T]^{-1}$

By Sherman morrison formula:

$$(A + UV^T)^{-1} = A^{-1} - \frac{A^{-1}U \cdot V^T A^{-1}}{1 + V^T A^{-1} U}$$

substitute $A = C \quad U = (-\bar{x}_i) \quad V^T = \bar{x}_i^T$

$$\therefore C_{(i)}^{-1} = C^{-1} - \frac{C^{-1}(-\bar{x}_i)(\bar{x}_i^T)C^{-1}}{1 + (\bar{x}_i^T)C^{-1}(-\bar{x}_i)}$$

$$\therefore \boxed{C_{(i)}^{-1} = C^{-1} + \frac{C^{-1}\bar{x}_i\bar{x}_i^T C^{-1}}{1 - \bar{x}_i^T C^{-1}\bar{x}_i}}$$

iv) Now, $\bar{\omega} = C^{-1}d$

$$\therefore \bar{\omega}_i = C_{(i)}^{-1} d_{(i)}$$

$$= \left[C^{-1} + \frac{C^{-1}\bar{x}_i\bar{x}_i^T C^{-1}}{1 - \bar{x}_i^T C^{-1}\bar{x}_i}\right][d - \bar{x}_i y_i]$$

$$\therefore \bar{\omega}_{(i)} = \underbrace{C^{-1} \cdot d}_{\bar{\omega}} + \frac{(C^{-1}\bar{x}_i\bar{x}_i^T C^{-1})d}{1 - \bar{x}_i^T C^{-1}\bar{x}_i} - (C^{-1})\cdot\bar{x}_i y_i$$

$$- \frac{(C^{-1})\bar{x}_i\bar{x}_i^T C^{-1}) \cdot \bar{x}_i y_i}{1 - \bar{x}_i^T C^{-1}\bar{x}_i}$$

$$\bar{\omega}_{(i)} = \bar{\omega} + (C^{-1}\bar{x}_i)\left[-y_i + \frac{\bar{x}_i^T \overbrace{(C^{-1} \cdot d)}^{\bar{\omega}}}{1 - \bar{x}_i^T C^{-1}\bar{x}_i} - \frac{\bar{x}_i^T C^{-1}\bar{x}_i y_i}{1 - \bar{x}_i^T C^{-1}\bar{x}_i}\right]$$

$$\therefore \bar{\omega}_{(i)} = \bar{\omega} + (C^{-1}\bar{x}_i)\left[-y_i + \frac{\bar{x}_i^T\bar{\omega} - \bar{x}_i^T C^{-1}\bar{x}_i y_i}{1 - \bar{x}_i^T C^{-1}\bar{x}_i}\right]$$

$$\boxed{\bar{\omega}_{(i)} = \bar{\omega} + (C^{-1}\bar{x}_i)\left[\frac{\cancel{-y_i + (\bar{x}_i^T C^{-1}\bar{x}_i)y_i} + \bar{x}_i^T\bar{\omega} - \cancel{\bar{x}_i^T C^{-1}\bar{x}_i y_i}}{1 - \bar{x}_i^T C^{-1}\bar{x}_i}\right]}$$

(y_i can be placed anywhere as it is scalar)

$$\therefore \quad \bar{w}_{(i)} = \bar{w} + (C^{-1}\bar{x}_i)\left[\frac{-y_i + \bar{x}_i^T \bar{w}}{1 - \bar{x}_i^T C^{-1} \bar{x}_i}\right]$$

v) $(\bar{w}^T \bar{x}_i - y_i) \Rightarrow$ To calculate this we take $\frac{1}{1 - \bar{x}_i^T C^{-1} \bar{x}_i}$ as λ because it is a scalar value

$$\therefore \quad \bar{w}^T\bar{x}_i - y_i = (\bar{w} + (C^{-1}\bar{x}_i)\lambda)^T \cdot \bar{x}_i - y_i$$

$$= (\bar{w}^T + \lambda^T (C^{-1} \cdot \bar{x}_i)^T)\, \bar{x}_i - y_i$$

$$= \bar{w}^T \bar{x}_i - y_i + \lambda^T (\bar{x}_i^T C^{-1T})\, \bar{x}_i$$

$$\Rightarrow \bar{w}^T \bar{x}_i - y_i + \left[\frac{-y_i + \bar{x}_i^T \bar{w}}{1 - \bar{x}_i^T C^{-1} \bar{x}_i}\right]^T (\bar{x}_i^T C^{-1T} \bar{x}_i)$$

$$= \bar{w}^T \bar{x}_i - y_i + \frac{-y_i + \bar{w}^T \bar{x}_i}{1 - \bar{x}_i^T C^{-1T} \bar{x}_i} (\bar{x}_i C^{-1T} \bar{x}_i)$$

$$= \frac{\bar{w}^T\bar{x}_i - (\bar{w}^T \bar{x}_i \bar{x}_i^T C^{-1T} \bar{x}_i) - y_i + y_i \bar{x}_i^T C^{-1T} \bar{x}_i + \bar{w}^T \bar{x}_i \bar{x}_i^T C^{-1T} \bar{x}_i - y_i \bar{x}_i^T C^{-1T} \bar{x}_i}{1 - \bar{x}_i^T C^{-1T} \bar{x}_i}$$

$$= \frac{\bar{w}^T \bar{x}_i - y_i}{(1 - \bar{x}_i^T C^{-1T} \bar{x}_i)} \rightarrow$$ (Taking transpose as it is scalar & T(scalar) = original)

$$\therefore \quad \bar{w}^T_{(i)} \bar{x}_i - y_i = \frac{\bar{w}^T \bar{x}_i - y_i}{1 - \bar{x}_i C^{-1} \bar{x}_i}$$

vi) LOOCV error $\Rightarrow \sum_{i=1}^{n} (\bar{\omega}_{(i)}^T \bar{x}_i - y_i)^2$

According to the formula in section 2.5

$$\bar{\omega}_{(i)}^T \cdot \bar{x}_i - y_i = \frac{\bar{\omega}^T \bar{x}_i - y_i}{1 - x_i^T C^{-1} x_i}$$

Here, we calculate C^{-1} only once, which has a complexity of $O(k^3)$

Multiplications require $O(k^2)$ time. [Additions & subtraction are linear]

We will perform these multiplications 'n' times.

$\therefore$ Total complexity $\Rightarrow O[k^3 + n(k^2+n)]$ [k+1 ≈ k for big k]

$\approx O[k^3]$

For usual method:

we will calculate $(C_{(i)})^{-1}$ everytime and use $\omega_{(i)} = (C_{(i)})^{-1} \cdot d_{(i)}$

This requires $O(n \times k^3)$ for all inverse calculation, and requires $O(n \times k^2)$ for multiplications as well.

$\therefore$ Total complexity would be $O[n(k^3 + nk^2)] \approx O[nk^3]$ which is higher than the previous one.