SML Assignment 2.

PRITISH WADHWA
2019440

A1) Derivation of MLE for bernoulli case.

I have to maximize $P(X|\theta)$.

now N iid samples

$$\therefore P(x_{id} | \theta_j) = \theta_j^{x_i} (1-\theta_j)^{1-x_i}.$$

$$\therefore \ln \prod_{i=1}^{n} P(x_i | \theta) = \ln \prod_{i=1}^{n} \prod_{j=1}^{d} P(x_{ij} | \theta_j)$$

$$= \ln \prod_{i=1}^{n} \prod_{j=1}^{d} (\theta_j^{x_{ij}})(1-\theta_j)^{1-x_{ij}}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{d} x_{ij} \ln \theta_j + (1-x_{ij}) \ln(1-\theta_j) = F(\theta)$$

$$\therefore \frac{\partial F(\theta)}{\partial \theta_j} = \sum_{i=1}^{n} \frac{x_{ij}}{\theta_j} - \frac{(1-x_{ij})}{1-\theta_j} = 0$$

$$\Rightarrow \sum_{i=1}^{n} \frac{(x_{ij} - \cancel{\theta_j x_{ij}} - \theta_j + \cancel{\theta_j x_{ij}})}{(\theta_j)(1-\theta_j)} = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_{ij} - n\theta_j = 0$$

$$\Rightarrow \theta_j \text{ or } \theta_{MLE} = \frac{\sum_{i=1}^{n} x_{ij}}{n}$$

(c) now Discriminant : $g_j(x) = \ln(p(x|\omega_j)) + \ln p(\omega_j)$

since the elements are statistically indep.

$$\therefore g_j(x) = \ln \prod_{i=1}^{d} p(x_i | \omega_j) + \ln p(\omega_j)$$

$$= \sum_{i=1}^{d} \ln p(x_i | \omega_j) + \ln p(\omega_j)$$

$$= \sum_{i=1}^{d} \ln(\theta_{ij}^{x_i} (1-\theta_{ij})^{(1-x_{ij})}) + \ln p(\omega_j)$$

$$= \sum_{i=1}^{d} x_i \ln \theta_{ij} + (1-x_{ij}) \ln(1-\theta_{ij})) + \ln p(\omega_j)$$

$$\therefore g_j(x) = \sum_{i=1}^{d} (x_i \ln \theta_{ij} + (1-x_i) \ln(1-\theta_{ij})) + \ln(p(\omega_j)).$$

$$= \sum_{i=1}^{d} x_i \ln \theta_{ij} + (1-x_i) \ln(1-\theta_{ij})) \quad (\ln p(\omega_j) \text{ same for both})$$

$\underline{\text{Ans 2)}}$ (a) Prior $= \theta_1 \theta_2 \ldots \theta_d e^{-(\theta_1 + \theta_2 + \ldots + \theta_d)}$

$$= \theta_1 \theta_2 \ldots \theta_d e^{-\theta_1} e^{-\theta_2} \ldots e^{-\theta_d}.$$

$$= (\theta_1 e^{-\theta_1})(\theta_2 e^{-\theta_2}) \ldots (\theta_d e^{-\theta_d})$$

$$= \prod_{i=1}^{d} \theta_i e^{-\theta_i}.$$

now $\theta_{MAP} = \arg\max_\theta P(D|\theta) P(\theta)$

$$= \arg\max_\theta \sum_{i=1}^{n} (\ln P(D|\theta)) + \ln P(\theta)$$

$$= \ln \prod_{i=1}^{n} \prod_{j=1}^{d} P(x_{ij}|\theta_j) + \ln P(\theta)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{d} (x_{ij} \ln \theta_j + (1-x_{ij}) \ln(1-\theta_j)) + \ln P(\theta)$$

now $\ln P(\theta) = \ln \prod_{i=1}^{d} \theta_i e^{-\theta_i}$

$$= \sum_{i=1}^{n} (\ln \theta_i - \theta_i)$$

$\therefore \theta_{map} = \arg\max_\theta \left(\sum_{i=1}^{n} \sum_{j=1}^{d} (x_{ij} \ln \theta_j + (1-x_{ij}) \ln(1-\theta_j)) + \sum_{j=1}^{n} (\ln \theta_j - \theta_j) \right)$

diff w.r.t θ_j & equate to 0.

$$\therefore \sum_{i=1}^{n} \left(\frac{x_{ij}}{\theta_j} - \frac{(1-x_{ij})}{1-\theta_j} \right) + \sum_{j=1}^{n} \left(\frac{1}{\theta_j} - 1 \right) = 0.$$

$$\Rightarrow \sum_{i=1}^{n} \left(\frac{x_{ij} - \cancel{\theta_j x_{ij}} - \theta_j + \cancel{\theta_j x_{ij}}}{(\theta_j)(1-\theta_j)} \right) \cancel{+ \sum_{j=1}^{n} \frac{(1-\theta_j)}{\theta_j}} = 0.$$

$$+ \frac{1 - \theta_j}{\theta_j} = 0$$

$$\Rightarrow \frac{\sum_{i=1}^{n} x_{ij} - n\theta_j}{(\theta_j)(1-\theta_j)} + \frac{(1-\theta_j)}{\theta_j} = 0.$$

$$\Rightarrow \sum_{i=1}^{n} x_{ij} - n\theta_j + (1-\theta_j)^2 = 0.$$

$$\Rightarrow \sum_{i=1}^{n} x_{ij} - n\theta_j + 1 + \theta_j^2 - 2\theta_j = 0.$$

$$\Rightarrow \theta_j^2 + \theta_j(-n-2) + \sum_{i=1}^{n} x_{ij} + 1 = 0.$$

$$\therefore \theta_j = \frac{(n+2) \pm \sqrt{(n+2)^2 - 4 - 4\sum_{i=1}^{n} x_{ij}}}{2} \qquad j \in \{1,2,\dots,d\}.$$

$$\therefore \theta_{MAP_j} = \frac{n+2 \pm \sqrt{n^2 + 4n - 4\sum_{i=1}^{n} x_{ij}}}{2} \qquad j \in \{1,2,\dots,d\}.$$

(b) Given Data :

$$X = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}_{d \times N}, \quad d = 2, \quad N = 4$$

$$\therefore \theta_{MAP_1} = \frac{(4+2) \pm \sqrt{4^2 + 4(4) - 4(1+0+1+1)}}{2}$$

$$= \frac{6 \pm \sqrt{32 - 12}}{2}$$

$$= \frac{6 \pm 2\sqrt{5}}{2}$$

$$\therefore \theta_{MAP_1} = 3 \pm \sqrt{5}$$

$$\theta_{MAP_2} = \frac{6 \pm \sqrt{32 - 4(1)}}{2}$$

$$= \frac{6 \pm 4\sqrt{7}}{2}$$

$$\theta_{MAP_2} = 3 \pm \sqrt{7}.$$

Ans3) (a) Chosen matrix

$$A = \begin{bmatrix} 2 & 6 \\ 5 & 1 \end{bmatrix}$$

$$\mu = \begin{bmatrix} (2+6)/2 \\ (5+1)/2 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$

$$A_c = A - \mu = \begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix}.$$

~~cov(Xme~~

$$\mathrm{cov}(A_c) = \left(\frac{1}{n-1}\right)(A_c)(A_c)^T$$

$$= \left(\frac{1}{1}\right)\begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix}\begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} 8 & -8 \\ -8 & 8 \end{bmatrix}$$

now to compute eigenvalues,

$$\begin{vmatrix} 8-\lambda & -8 \\ -8 & 8-\lambda \end{vmatrix} = 0.$$

$$\Rightarrow (8-\lambda)^2 - (+64) = 0$$

$$\Rightarrow 64 + \lambda^2 - 16\lambda - 64 = 0$$

$$\Rightarrow (\lambda)(\lambda - 16) = 0 \Rightarrow \left.\begin{array}{l} \lambda = 0 \\ \lambda = 16 \end{array}\right\} \text{eigenvalues}$$

$\therefore$ Eigenvector corresponding to $\lambda = 16$:

$$M: \begin{bmatrix} 8-\lambda & -8 \\ -8 & 8-\lambda \end{bmatrix} = \begin{bmatrix} -8 & -8 \\ -8 & -8 \end{bmatrix} \xrightarrow[\text{RREF}]{\text{convert to}} \xrightarrow{R_1 \to R_1/-8} \begin{bmatrix} 1 & 1 \\ -8 & -8 \end{bmatrix}$$

$$\downarrow R_2 \to R_2 + 8R_1$$

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}.$$

$$\therefore \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow x_1 + x_2 = 0 \Rightarrow x_1 = -x_2.$$

$\therefore$ Eigenvector corresponding to $\lambda = 16$: $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$

normalized eigenvector corresponding to $\lambda = 16$ $= \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$

Eigenvector corresponding to $\lambda = 0$.

$$\begin{bmatrix} 8-\lambda & -8 \\ -8 & 8-\lambda \end{bmatrix} = \begin{bmatrix} 8 & -8 \\ -8 & 8 \end{bmatrix} \xrightarrow[\text{to RREF}]{\text{Convert}} \begin{cases} R_1 \to R_1/8 \end{cases}$$

$$\begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \xleftarrow{R_2 \to R_2 + 8R_1} \begin{bmatrix} 1 & -1 \\ -8 & 8 \end{bmatrix}$$

$$\therefore \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$\Rightarrow \quad x_1 - x_2 = 0.$

$\Rightarrow \quad x_1 = x_2.$

$\therefore$ Eigenvector corresponding to $\lambda = 0$: $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

normalized Eigenvector corresponding to $\lambda = 0$: $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$.

$$\therefore \quad U = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

$$Y = U^T A_c.$$

$$= \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix} = \begin{bmatrix} -2\sqrt{2} & 2\sqrt{2} \\ 0 & 0 \end{bmatrix}$$

(b). $UY + \text{mean}(X)$

$$= \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} -2\sqrt{2} & 2\sqrt{2} \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$ { row wise ~~sum~~ element sum.

$$= \begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix} + \begin{bmatrix} 4 \\ 3 \end{bmatrix} = \begin{bmatrix} 2 & 6 \\ 5 & 1 \end{bmatrix}$$

$MSE\left(UY + \text{mean}(X), X\right) = 0$ as all the elements are same in $UY + \text{mean}(X)$ and X.

Scanned with CamScanner