___ / / ___

SML - Assign 2

Q3. (a) Let the arbitrary matrix be

$$X = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}_{2\times 2}$$

$$\mu = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$X - \mu = X_c = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

$$S_{x_c} = \frac{1}{2} \sum_{i=1}^{2} x_i x_i^T = \frac{1}{2}\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \end{bmatrix} \right)$$

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

$$= \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

To compute eigen values of S_{x_c},

$$S_{x_c} v_1 = \lambda u_1$$

P.T.O

$\Rightarrow \quad (S_{x_c} - \lambda I)\, u_1 = 0.$

$\Rightarrow \quad \det(S_{x_c} - \lambda I) = 0.$

$$\det \begin{bmatrix} 1-\lambda & -1 \\ -1 & 1-\lambda \end{bmatrix} = 0.$$

$\Rightarrow \quad (1-\lambda)^2 - 1 = 0$

$\Rightarrow \quad 1 + \lambda^2 - 2\lambda - 1 = 0.$

$\Rightarrow \quad \lambda^2 - 2\lambda = 0.$

$\Rightarrow \quad \lambda(\lambda - 2) = 0$

$\Rightarrow \quad \underline{\lambda = 0} \quad \text{or} \quad \underline{\lambda = 2}$

Case I. $\lambda = 0$,

$$S_{x_c} - \lambda I = \begin{bmatrix} +1 & -1 \\ -1 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \underbrace{\begin{bmatrix} u_{11} \\ u_{12} \end{bmatrix}}_{u_1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

$$\left[\begin{array}{cc|c} 1 & -1 & 0 \\ -1 & 1 & 0 \end{array}\right] \rightarrow \left[\begin{array}{cc|c} 1 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

$\Rightarrow \quad u_{11} = u_{12}$

P.T.O.

$\therefore \quad u_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ as $\underline{u_1^T u_1 = 1}$

Case II $\quad \lambda = 2,$

$$S_{x_c} - \lambda I = \begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix}$$

$$\begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix} \begin{array}{c} u_2 \\ \downarrow \\ \begin{bmatrix} u_{21} \\ u_{22} \end{bmatrix} \end{array} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\left[\begin{array}{cc|c} -1 & -1 & 0 \\ -1 & -1 & 0 \end{array}\right] \rightarrow \left[\begin{array}{cc|c} -1 & -1 & 0 \\ 0 & 0 & 0 \end{array}\right]$$

$\therefore \quad u_1 = -u_2 \quad \Rightarrow$

$$u_2 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$$

$\therefore \quad U = \begin{bmatrix} u_2 & u_1 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$ = PCA Matrix.

$\because$ Sorting according to decreasing order of eigenvalues.

$$Y = U^T x_c = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

PTO.

$$\Rightarrow \quad Y = \begin{bmatrix} \sqrt{2} & -\sqrt{2} \\ 0 & 0 \end{bmatrix}$$

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

$$= U U^T X_c + \text{mean}(X)$$

$$= X_c + \text{mean}(X)$$

$$= \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$X = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$MSE = \frac{(2-2)^2 + (0-0)^2 + (0-0)^2 + (2-2)^2}{4}$$

$$= \boxed{0}$$

(c) Yes, the calculation matches the code

(d) $X = MVG(\mu, \Sigma)$ $\quad \mu \to d\times 1$

$\Sigma \to d\times d$

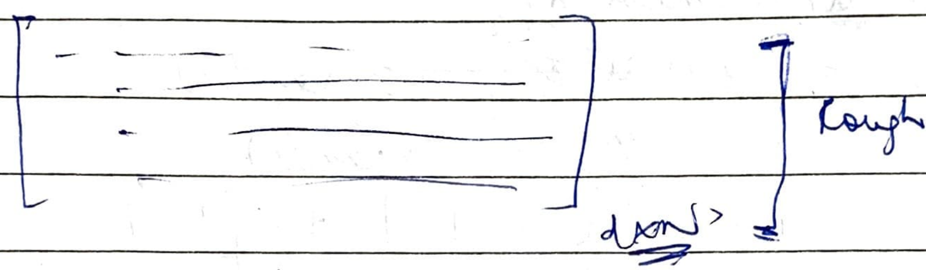
N samples in total

1.1.0

$x \to$ ~~[scribble]~~ $\mu_{calc}(x) \to x - \mu_{calc}(x) = x_c$

$S_{x_c} \to u \to r \to$ MSE ← (plan)

d) $x_c \to$



I plotted this using scipy, multivariate_normal & sklearn.datasets

(e.) Refer to code.

(f). Refer to code.
The value for MSE comes out to be 0 as there is no data loss.

(g) In this case, there is always some data lost in the samples. Therefore, MSE $\neq 0$ unless $p = d$.

Q1. (a) Done using ~~(add)~~ scipy. bernoulli

(b) Done using MLE.

* Multivariate bernoulli

θ_1 , θ_2

θ_{ij} 'j' denotes dimension

$$\theta_j = Pr(x_j = 1 \mid \theta_j)$$

N samples

$$Pr(x_j \mid \theta_j) = \theta_j^{x_j} (1-\theta_j)^{1-x_j}$$

$$Pr(x \mid \theta) = \prod_{j=1}^{2} \theta_j^{x_j} (1-\theta_j)^{1-x_j}$$

(Assuming independence)

.) For N iid samples,

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

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

P.T.O

/ /

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

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

$$\Rightarrow \boxed{\theta_j = \frac{1}{N} \sum_{i=1}^{N} x_{ij}}$$

$$\Rightarrow \theta_1 = \frac{1}{N} \sum_{i=1}^{N} x_{i1}$$

$$\theta_2 = \frac{1}{N} \sum_{i=1}^{N} x_{i2}$$

For our case

•) As the number of samples increase, the readings become more & more accurate in case of MLE.

$\Rightarrow$ [illegible] $\frac{\sum_{i=1}^{50} x_i}{50}$ gives most accurate data in our case.

Q2. (a.) $P(\theta) = \theta_1 \theta_2 \theta_3 \dots \theta_d \, e^{\theta_1 + \theta_2 + \theta_3 \dots \theta_d}$

For θ_{MAP},

$$\max_{\theta} \left[\sum_{i=1}^{N} P(x|\theta) \right] + \ln P(\theta)$$

$$F(\theta) = \sum_{i=1}^{N} \sum_{j=1}^{d} \left[x_{ij} \ln \theta_j + (1-x_{ij})(1-\theta_j) \right]$$

$$+ \ln P(\theta)$$

(Assuming independence)

to max, $\frac{\partial . F(\theta)}{\partial \theta_j} = 0$

$$\Rightarrow \sum_{i=1}^{N} \left[\frac{x_{ij}}{\theta_j} - \frac{(1-x_{ij})}{(1-\theta_j)} \right] + \frac{\partial}{\partial \theta_j} \ln \theta_1 \theta_2 \dots \theta_j \dots \theta_d \, e^{(\theta_1 + \theta_2 \dots \theta_j \dots \theta_d)} = 0.$$

$$\Rightarrow \sum_{i=1}^{N} \left[\frac{x_{ij}}{\theta_j} - \frac{(1-x_{ij})}{(1-\theta_j)} \right] + \frac{1}{(\theta_1 \theta_2 \dots \theta_j \dots \theta_d) e^{(\theta_1 + \theta_2 \dots \theta_j \dots + \theta_d)}} \cdot \frac{\partial . P'(\theta)}{\partial \theta_j}$$

P.T.O.

$$\Rightarrow \sum_{i=1}^{N} \frac{x_{ij}}{\theta_j} - \frac{(1-x_{ij})}{(1-\theta_j)} + \frac{1}{P(\theta)} \left[\theta_1 \theta_2 \ldots \theta_{j-1} \theta_{j+1} \ldots \theta d \, e^{-\theta_1 + \theta_2 \ldots \theta_j + \ldots \theta d} \right.$$

$$\left. + (\theta_1 \theta_2 \ldots \theta_j \ldots \theta d) e^{\theta_1 + \theta_2 \ldots \theta_j + \ldots \theta d} \cdot \cancel{(1)} \right] = 0$$

$$\Rightarrow \sum_{i=1} \frac{x_{ij}}{\theta_j} - \frac{(1-x_{ij})}{1-\theta_j} + \frac{1}{\cancel{P(\theta)}} \cdot \left(\frac{\cancel{P(\theta)}}{\theta_j} - \cancel{P(\theta)} \right) = 0.$$

$$\Rightarrow \sum_{i=1}^{N} \left[\frac{x_{ij}}{\theta_j} - \frac{(1-x_{ij})}{(1-\theta_j)} \right] + \frac{1-\theta_j}{\theta_j} = 0.$$

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

$$\Rightarrow \frac{(1-\theta_j) \sum_{i=1}^{N} x_{ij}}{\cancel{(1-\theta_j)} \cancel{\theta_j}} - \frac{\theta_j}{\cancel{\theta_j} \cancel{(1-\theta_j)}} \left[N - \sum_{i=1}^{N} x_{ij} \right] + \frac{(1-\theta_j)^2}{\cancel{\theta_j} \cancel{(1-\theta_j)}} = 0$$

$$\Rightarrow \sum x_{ij} - \cancel{\theta_j \sum x_{ij}} - \theta_j N + \cancel{\theta_j \sum x_{ij}} + (1-\theta_j)^2 = 0$$

$$\theta_j^2 + 1 - 2\theta_j - \theta_j N + \sum x_{ij} = 0.$$

$$\theta_j = \frac{+(N+2) \pm \sqrt{(N+2)^2 - 4\left(1 + \sum_{i=1}^{N} x_{ij}\right)}}{2}$$

$$\Rightarrow \boxed{\theta_j = \frac{(N+2) \pm \sqrt{N^2 + 4N - 4\sum_{i=1}^{N} x_{ij}}}{2}}$$

(b). $x = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$ $d = 2$, $N = 4$.

$$\theta_1 = \frac{6 \pm \sqrt{16 + 16 - 4(3)}}{2} = \frac{6 \pm 2\sqrt{5}}{2} = \boxed{3 \pm \sqrt{5}}$$

PTO.

As $0 \leq \theta_j \leq 1$, $\boxed{\theta_1 = 3 - \sqrt{5}}$

$$\theta_2 = \frac{6 \pm \sqrt{16 + 16 - 4(1)}}{2}$$

$$= \frac{6 \pm 2\sqrt{7}}{2} = 3 \pm \sqrt{7}.$$

$\Rightarrow \boxed{\theta_2 = 3 - \sqrt{7}}$ //

$\therefore \boxed{\theta_{MAP} = \begin{bmatrix} 3 - \sqrt{5} \\ 3 - \sqrt{7} \end{bmatrix}}$ //

/ /

Q1. (e) $\omega_1, \omega_2 \cdots \omega_c$

$g_i(x) \rightarrow \forall i = 1, \ldots, c$

$$P(x|\omega_1) = \prod_{i=1}^{d} P_i^{x_i} (1-P_i)^{1-x_i}$$

$$P_{ij} = P_r(x_i = 1 | \omega_j)$$

$$P_{ij} = P_r(x_i = 1 | \omega_1) \qquad 1 - P_{i1} = 1 - P_r(x_i = 1 | \omega_1)$$

$$P(x|\omega_j) = \prod_{i=1}^{d} P_{ij}^{x_j} (1-P_{ij})^{1-x_i}$$

$$P(x_i|\omega_j) = P_{ij}^{x_{ij}} (1-P_{ij})^{1-x_{ij}}$$

$$g_j(x) = \ln P(x|\omega_j) + \ln P(\omega_j)$$

$\forall j = 1, 2, \ldots, c$

$\max_j g_j(x) \rightarrow$ map x to j.

$$g_j(x) = \ln \prod_{i=1}^{N} \prod_{j=1}^{d} P_{ij}^{x_{ij}} (1-P_{ij})^{1-x_{ij}} + \ln P(\omega_j)$$

P.T.O

As we only have 2 classes to classify,

$$= \sum_{i=1}^{N}\sum_{j=1}^{d} x_{ij} \ln P_{ij} + (1-x_{ij}) \ln (1-P_{ij}) + \ln P(\omega_j).$$

$$g_1(x) = \sum_{i=1}^{N}\sum_{j=1}^{d} x_{ij} \ln P_{ij} + (1-x_{ij}) \ln (1-P_{ij}) + \ln (\omega_1)$$

$$g_2(x) = \sum_{i=1}^{N}\sum_{j=1}^{d} x_{ij} \ln P_{i2} + (1-x_{ij}) \ln (1-P_{i2}) + \ln (\omega_2)$$

Here, we has assumed the priors are equal

$$\therefore \boxed{\underbrace{g(x)}_{\text{dichotomizer}} = \sum_{j=1}^{d} x_{ij} \ln P_{i1} + (1-x_{ij}) \ln (1-P_{i1}) - x_i \ln P_{i2} - (1-x_{ij}) \ln (1-P_{i2})}$$

$$g(x) = \sum_{i=1}^{N}\sum_{j=1}^{d} x_i \frac{\ln P_{i1}}{\ln P_{i2}} + (1-x_{ij}) \ln \frac{(1-P_{i1})}{(1-P_{i2})}$$

In our case,

$P_{11} = 0.5 \qquad P_{21} = 0.8$

$P_{12} = 0.9 \qquad P_{22} = 0.2$

$$g(x) = x_1 \ln\left(\frac{0.5}{0.9}\right) + (1-x_1)\ln\left(\frac{0.5}{0.1}\right)$$

upon solving

$$+ x_2 \ln\left(\frac{0.8}{0.2}\right) + (1-x_2)\ln\left(\frac{0.2}{0.8}\right)$$

$$\boxed{g(x) = x_2 \ln 16 - x_1 \ln 9 + \ln\left(\frac{5}{4}\right)}$$