Q1) Given data matrix $X \in R^{d \times N}$,

the projection of any $x \in X$ is $U_1^T x$

where U_1 is a column of Covariance Matrix.

We want to maximise variance such that the projected mean is 1.

$Y = U_1^T X$

$Var(y) = \frac{1}{N-1} \sum \| y_i - mean(y) \|^2$

constrained on $mean(y) = \| U^T \mu \|^2 = 1$, where $\mu = mean(X)$

Using Lagrange multipliers,

$L = \frac{1}{N-1} \sum \| U^T x_i - U^T \mu \|^2 - \lambda (\| U^T \mu \|^2 - 1)$ 1.5

$\frac{\nabla L}{\nabla U} = \frac{2}{N-1} \left[\sum (x_i - \mu)(x_i - \mu)^T \right] U - 2\lambda \mu \mu^T U = 0$

$= 2CU - 2\lambda \mu \mu^T U = 0$ 1.5

$\Rightarrow CU = \lambda \mu \mu^T U$

Solve for eigen vector and we will get projection vector with largest eigen value.

2.) Given $d(\mu_1, \mu_2) = 1$

$$\bar{\mu}_1 = w^T \mu_1$$

$$\bar{\mu}_2 = w^T \mu_2$$

$$d(\bar{\mu}_1, \bar{\mu}_2) = (w^T \mu_1 - w^T \mu_2)(w^T \mu_1 - w^T \mu_2)^T$$

$$= w^T \underbrace{(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T}_{S_b} w$$

$w^T S_b w = 1$

We maximize Total scatter $S = \sum_{i=1}^{2N} (x_i - \mu)(x_i - \mu)^T$

μ is the global mean $= \frac{\mu_1 + \mu_2}{2}$

$x_1 \in R^{d \times N}$ $x_2 \in R^{d \times N}$, so total $2N$ samples

Also we minimize the within ~~scat~~ class scatter $w^T S_w w$ where $S_w = S_1 + S_2$

$S_1 = N \cdot \Sigma_1$ $\quad S_2 = N \Sigma_2$

Our objective function is

$$\max_{\omega} \frac{\omega^T S \omega}{\omega^T S_W \omega}$$

1.5 $$\max \ \omega^T S \omega \quad \text{s.t. } \omega^T S_B \omega = 1 \ \& \ \omega^T S_W \omega = 1$$

$$\frac{\partial}{\partial \omega}\left(\omega^T S \omega - \lambda_1\left(\omega^T S_B \omega - 1\right) - \lambda_2\left(\omega^T S_W \omega - 1\right)\right) = 0$$

$$\Rightarrow S\omega - \lambda_1 S_B \omega - \lambda_2 S_W \omega = 0$$

1.5 $$\Rightarrow S \omega = \left(\lambda_1 S_B + \lambda_2 S_W\right) \omega$$

Sol^n ③ Let the discriminant function be

$$g_i(x) = \ln p(\omega_i | x)$$

$$= \ln\left(\frac{p(x|\omega_i)\, p(\omega_i)}{p(x)} \right)$$

$$\equiv \ln p(x|\omega_i) + \ln p(\omega_i)$$

Given,

$$p(x|\omega_1) = \prod_{i=1}^{d} \theta_i^{x_i} (1-\theta_i)^{1-x_i} \, \frac{\theta_i^{\alpha_i - 1} (1-\theta_i)^{\beta_i - 1}}{Z_i}$$

$$\ln p(x|\omega_1) = \ln\left(\prod_{i=1}^{d} \theta_i^{x_i} (1-\theta_i)^{1-x_i} \, \frac{\theta_i^{\alpha_i - 1} (1-\theta_i)^{\beta_i - 1}}{Z_i} \right)$$

$$= \sum_{i=1}^{d} \Big[x_i \ln\theta_i + (1-x_i)\ln(1-\theta_i) + (\alpha_i - 1)\ln\theta_i + (\beta_i - 1)\,(1-\theta_i) - \ln Z_i \Big]$$

$$= \sum_{i=1}^{d} \Big[x_i(\ln\theta_i - \ln(1-\theta_i)) + \cancel{\ln(1-\theta_i)} + (\alpha_i - 1)\ln\theta_i + \beta_i \ln(1-\theta_i) - \cancel{\ln(1-\theta_i)} - \ln Z_i \Big]$$

1.5 $$= \sum_{i=1}^{d} \left[x_i \ln\left(\frac{\theta_i}{1-\theta_i}\right) + (\alpha_i - 1)\ln\theta_i + \beta_i \ln(1-\theta_i) - \ln Z_i \right]$$

$$= \sum_{i=1}^{d} x_i \ln\left(\frac{\theta_i}{1-\theta_i}\right) + \sum_{i=1}^{d}\left[(\alpha_i - 1)\ln\theta_i + \beta_i \ln(1-\theta_i) - \ln z_i\right]$$

$$= w_\theta^T x + \sum_{i=1}^{d}\left[(\alpha_i - 1)\ln\theta_i + \beta_i \ln(1-\theta_i) - \ln z_i\right]$$

where $w_\theta = \begin{bmatrix} \ln\left(\frac{\theta_1}{1-\theta_1}\right) \\ \ln\left(\frac{\theta_2}{1-\theta_2}\right) \\ \vdots \\ \ln\left(\frac{\theta_d}{1-\theta_d}\right) \end{bmatrix}$ and $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$

Similarly,

$$\ln p(x|\omega_2) = w_\phi^T x + \sum_{i=1}^{d}\left[(\alpha_i - 1)\ln\phi_i + \beta_i \ln(1-\phi_i) - \ln z_i\right]$$

where $w_\phi = \begin{bmatrix} \ln\left(\frac{\phi_1}{1-\phi_1}\right) \\ \ln\left(\frac{\phi_2}{1-\phi_2}\right) \\ \vdots \\ \ln\left(\frac{\phi_d}{1-\phi_d}\right) \end{bmatrix}$

Decision Boundary

$g_1(x) = g_2(x)$

$\Rightarrow \ln p(x|\omega_1) + \ln p(\omega_1) = \ln p(x|\omega_2) + \ln p(\omega_2)$

$$\Rightarrow w_\theta^T x + \sum_{i=1}^{d} \left[(\alpha_i - 1)\ln\theta_i + \beta_i \ln(1-\theta_i) - \cancel{\ln z_i} \right] + \ln p(\omega_1)$$

$$= w_\phi^T x + \sum_{i=1}^{d} \left[(\alpha_i - 1)\ln\phi_i + \beta_i \ln(1-\phi_i) - \cancel{\ln z_i} \right] + \ln p(\omega_2)$$

$$\Rightarrow (w_\theta - w_\phi)^T x + \sum_{i=1}^{d} \left[(\alpha_i - 1)\ln\left(\frac{\theta_i}{\phi_i}\right) + \beta_i \ln\frac{(1-\theta_i)}{(1-\phi_i)} \right] + \ln\frac{p(\omega_1)}{p(\omega_2)} = 0$$

Text

which is of the form

$$w^T x + b = 0$$

where $w = w_\theta - w_\phi$

1.5

$$= \begin{bmatrix} \ln\left(\frac{\theta_1(1-\phi_1)}{\phi_1(1-\theta_1)}\right) \\ \ln\left(\frac{\theta_2(1-\phi_2)}{\phi_2(1-\theta_2)}\right) \\ \vdots \\ \ln\left(\frac{\theta_d(1-\phi_d)}{\phi_d(1-_d)}\right) \end{bmatrix}$$

and $b = \sum_{i=1}^{d} \left[(\alpha_i - 1)\ln\left(\frac{\theta_i}{\phi_i}\right) + \beta_i \ln\left(\frac{1-\theta_i}{1-\phi_i}\right) \right] + \ln\left(\frac{p(\omega_1)}{p(\omega_2)}\right)$

Q4.) $X = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix} \in \mathbb{R}^{2\times 2}$ $d=2$, $N=2$.

$$\mu = \begin{pmatrix} \frac{2+0}{2} \\ \frac{0-2}{2} \end{pmatrix} = \begin{pmatrix} \frac{2}{2} \\ \frac{-2}{2} \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Now centralize X along μ : $X \leftarrow X - \mu$

$$X - \mu = \begin{pmatrix} 2-1 & 0-1 \\ 0-(-1) & -2-(-1) \end{pmatrix} = \begin{pmatrix} 2-1 & -1 \\ 1 & -2+1 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}.$$

$$\Sigma = Cov(X) = \underbrace{\frac{1}{N-1}}_{\text{given, use unbiased estimator}} XX^T = 1 \cdot \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix}.$$

$$= \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}$$ 1

Now, $\det(\Sigma - \lambda I) = 0 \Rightarrow \begin{vmatrix} 2-\lambda & 2 \\ 2 & 2-\lambda \end{vmatrix} = 0$

$\Rightarrow (2-\lambda)(2-\lambda) - 4 = 0 \Rightarrow (\lambda-2)^2 = 4$

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

$\Rightarrow (\lambda-4)(\lambda) = 0$

$\Rightarrow \lambda_1 = 4$ and $\lambda_2 = 0$ 1

<u>Case 1</u>: For $\lambda_1 = 4$ we find u_1 : $\Sigma u_1 = \lambda_1 u_1$

$\Rightarrow \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 4 \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$

$\Rightarrow 2x_1 + 2x_2 = 4x_1 \Rightarrow 2x_1 = 2x_2 \Rightarrow x_1 = x_2.$

$2x_1 + 2x_2 = 4x_2 \Rightarrow 2x_1 = 2x_2 \Rightarrow u_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$

<u>Case 2</u>: For $\lambda_2 = 0$ we find u_2 : $\Sigma u_2 = \lambda_2 u_2$ 1

$\Rightarrow \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0 \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \Rightarrow \begin{matrix} 2x_1 + 2x_2 = 0 \\ 2x_1 + 2x_2 = 0 \end{matrix} \Rightarrow \begin{matrix} x_1 = -x_2 \\ u_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}. \end{matrix}$

$\therefore$ PCA matrix is $[u_1 \; u_2] = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$.

Other answers where $u_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ are also valid.

Q5. We have $p(x|\theta) = \prod_{i=1}^{d} \theta_i^{x_i} (1-\theta_i)^{1-x_i}$

Prior: $P(\theta) = \prod_{i=1}^{d} P(\theta_i)$ [assuming iid for θ_i]

$$= \prod_{i=1}^{d} \left(\frac{\theta_i^{\alpha_i - 1} (1-\theta_i)^{\beta_i - 1}}{Z_i} \right)$$

For n observations, likelihood:

$$P(X|\theta) = \prod_{j=1}^{n} p(x^j|\theta)$$

$$= \prod_{j=1}^{n} \prod_{i=1}^{d} \theta_i^{x_i^j} (1-\theta_i)^{(1-x_i^j)}$$

where x^j represents the j^{th} observation in X.

MAP estimate: $\hat{\theta}_{i_{MAP}} = \underset{\theta_i}{\operatorname{argmax}} \; P(X|\theta) \cdot P(\theta)$

$$= \underset{\theta_i}{\operatorname{argmax}} \left[\prod_{j=1}^{n} \prod_{i=1}^{d} \theta_i^{x_i^j} (1-\theta_i)^{(1-x_i^j)} \right] \cdot \left(\prod_{i=1}^{d} \frac{\theta_i^{\alpha_i - 1} (1-\theta_i)^{\beta_i - 1}}{Z_i} \right)$$

$$\hat{\theta}_{i_{MAP}} = \underset{\theta_i}{\operatorname{argmax}} \log \left(\left(\prod_{j=1}^{n} \prod_{i=1}^{d} \theta_i^{x_i^j} (1-\theta_i)^{(1-x_i^j)} \right) \cdot \left(\prod_{i=1}^{d} \frac{\theta_i^{\alpha_i - 1} (1-\theta_i)^{\beta_i - 1}}{Z_i} \right) \right)$$

1 $$= \underset{\theta_i}{\operatorname{argmax}} \left(\sum_{j=1}^{n} \sum_{i=1}^{d} \left(x_i^j \log \theta_i + (1-x_i^j) \log(1-\theta_i) \right) \right)$$

$$+ \sum_{i=1}^{d} \left((\alpha_i - 1) \log \theta_i + (\beta_i - 1) \log(1-\theta_i) - \log Z_i \right)$$

Taking partial derivative wrt θ_i and equating to 0.
(All $i \neq i$ will ~~go~~ go to zero)

$$\sum_{j=1}^{n} \left(\frac{x_i^j}{\theta_i} - \frac{1-x_i^j}{(1-\theta_i)} \right) + \frac{\alpha_i - 1}{\theta_i} - \frac{\beta_i - 1}{1-\theta_i} = 0$$

1

$$\sum_{j=1}^{n} \left(\frac{x_i^j - \cancel{\theta x_i^j} - \theta_i + \cancel{x_i^j \theta_i}}{\theta_i (1-\theta_i)} \right)$$

$$+ \frac{\alpha_i - \alpha_i \theta_i - 1 + \theta_i - \beta_i \theta_i + \theta_i}{\theta_i (1-\theta_i)} = 0$$

$$\sum_{j=1}^{n} \left(x_i^j - \theta_i \right) + (\alpha_i - 1) + \theta_i (2 - \alpha_i - \beta_i) = 0$$

$$\sum_{j=1}^{n} x_i^j - n\theta_i + \alpha_i - 1 - \theta_i (\alpha_i + \beta_i - 2) = 0$$

$$\theta_i (n + \alpha_i + \beta_i - 2) = \left(\sum_{j=1}^{n} x_i^j \right) + \alpha_i - 1$$

$$\boxed{\hat{\theta}_{i_{MAP}} = \theta_i = \frac{\left(\sum_{j=1}^{n} x_i^j \right) + \alpha_i - 1}{n + \alpha_i + \beta_i - 2}}$$ 1