HW2

Q1-

$$\mu = \frac{\begin{bmatrix}3\\3\end{bmatrix}+\begin{bmatrix}-3\\-3\end{bmatrix}}{2} = 0$$

Part A)

$$\Sigma_B = \sum_{c=1}^{2}(\mu_c - \mu)(\mu_c - \mu)^T = \begin{bmatrix}3\\3\end{bmatrix}\begin{bmatrix}3 & 3\end{bmatrix} + \begin{bmatrix}-3\\-3\end{bmatrix}\begin{bmatrix}-3 & -3\end{bmatrix} = 18\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}$$

Part B)

$$\Sigma = \Sigma_B + \Sigma_W = 18\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix} + \begin{bmatrix}4 & 4\\4 & 9\end{bmatrix} = \begin{bmatrix}22 & 22\\22 & 27\end{bmatrix}$$

$$\Sigma^{-1}\Sigma_B = \frac{18}{22\times 5}\begin{bmatrix}27 & -22\\-22 & 22\end{bmatrix}\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix} = \frac{9}{55}\begin{bmatrix}5 & 5\\0 & 0\end{bmatrix} = \begin{bmatrix}\frac{9}{11} & \frac{9}{11}\\0 & 0\end{bmatrix} \qquad \Lambda = \begin{bmatrix}\frac{9}{11} & 0\\0 & 0\end{bmatrix} \qquad \Phi = \begin{bmatrix}1 & \frac{1}{\sqrt{2}}\\0 & \frac{-1}{\sqrt{2}}\end{bmatrix}$$

Part C)

$$\Phi^T\Sigma_B\Phi = 18\begin{bmatrix}1 & 0\\\frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}\begin{bmatrix}1 & \frac{1}{\sqrt{2}}\\0 & \frac{-1}{\sqrt{2}}\end{bmatrix} = \begin{bmatrix}1 & 1\\0 & 0\end{bmatrix}\begin{bmatrix}1 & \frac{1}{\sqrt{2}}\\0 & \frac{-1}{\sqrt{2}}\end{bmatrix} = \begin{bmatrix}18 & 0\\0 & 0\end{bmatrix}$$

$$\Phi^T\Sigma\Phi = \begin{bmatrix}1 & 0\\\frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}22 & 22\\22 & 27\end{bmatrix}\begin{bmatrix}1 & \frac{1}{\sqrt{2}}\\0 & \frac{-1}{\sqrt{2}}\end{bmatrix} = \begin{bmatrix}22 & 22\\0 & \frac{-5}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}1 & \frac{1}{\sqrt{2}}\\0 & \frac{-1}{\sqrt{2}}\end{bmatrix} = \begin{bmatrix}22 & 0\\0 & 2.5\end{bmatrix}$$

Part D)

How we arrange the eigenvector matrix affects the output. Considering $\Phi = \begin{bmatrix}1 & \frac{1}{\sqrt{2}}\\0 & \frac{-1}{\sqrt{2}}\end{bmatrix}$ leads to image 2 and $\Phi = \begin{bmatrix}\frac{1}{\sqrt{2}} & 1\\\frac{-1}{\sqrt{2}} & 0\end{bmatrix}$ leads to image 3.

Q2-

Both eigenvalues are 1 and eigenvectors are $\begin{bmatrix}0\\1\end{bmatrix}$ and $\begin{bmatrix}0\\-1\end{bmatrix}$. So, the matrix $\begin{bmatrix}1 & 0\\1 & 1\end{bmatrix}$ does not have two linearly independent eigenvectors and this is why it is not diagonalizable. This is not symmetric and we do not get a unique value for ρ.

Q5-

$S = m_1 m_1^T + m_2 m_2^T$

$\sum \lambda_i = tr(S) = tr(m_1 m_1^T) + tr(m_2 m_2^T) = tr(m_1^T m_1) + tr(m_2^T m_2) = \|m_1\|^2 + \|m_2\|^2 = 5$

Rank of S is 2. Therefore, we have 2 eigenvalues: $\lambda_1 + \lambda_2 = 5$.

The angle between m_1 and m_2 is 60 degrees and a plane passes through these vectors. Considering $S\Phi = \lambda\Phi$, Φ in this plane can be written in terms of these vectors: $\Phi = m_1 + \alpha m_2$

$S\Phi = \lambda\Phi \rightarrow (m_1 m_1^T + m_2 m_2^T)(m_1 + \alpha m_2) = \lambda(m_1 + \alpha m_2) \rightarrow m_1 m_1^T m_1 + m_2 m_2^T m_1 + \alpha m_1 m_1^T m_2 + \alpha m_2 m_2^T m_2 = \lambda m_1 + \lambda\alpha m_2$

$m_1^T m_1 = \|m_1\|^2 = 1 \quad m_2^T m_2 = \|m_2\|^2 = 4 \quad m_1^T m_2 = m_2^T m_1 = \|m_1\|\|m_2\| \cos(60) = 1 \times 2 \times \frac{1}{2} = 1$

$S\Phi = \lambda\Phi \rightarrow m_1 + m_2 + \alpha m_1 + 4\alpha m_2 = (1+\alpha)m_1 + (1+4\alpha)m_2 = \lambda m_1 + \lambda\alpha m_2$

$1 + \alpha = \lambda$ and $1 + 4\alpha = \lambda\alpha$ $\rightarrow$ 2 equations, 2 unknowns $\rightarrow$ $\lambda_1, \lambda_2 = \frac{5 \pm \sqrt{13}}{2}$

Q7-

Part A) $\frac{\partial}{\partial m} \sum_{i=1}^{N} (x_i - m)^T (x_i - m) = 0 \rightarrow m = \frac{1}{N} \sum_{i=1}^{N} x_i$

Part B) $\frac{\partial}{\partial \alpha_i} \sum_{i=1}^{N} (x_i - m - \alpha_i e)^T (x_i - m - \alpha_i e) = 0 \rightarrow e^T (x_i - m - \alpha_i e) = 0 \rightarrow \alpha_i = e^T (x_i - m)$

Part C) $\frac{\partial}{\partial e} \sum_{i=1}^{N} (x_i - m - \alpha_i e)^T (x_i - m - \alpha_i e) = 0 \rightarrow \sum_{i=1}^{N} \alpha_i (x_i - m) = \sum_{i=1}^{N} \alpha_i^2 e$

$\sum_{i=1}^{N} (x_i - m)(x_i - m)^T e = \sum_{i=1}^{N} \alpha_i^2 e \rightarrow \frac{1}{N} \sum_{i=1}^{N} (x_i - m)(x_i - m)^T e = \frac{1}{N} \sum_{i=1}^{N} \alpha_i^2 e \rightarrow Se = \lambda e$

e is the eigenvector of S and it is the eigenvector corresponding to greater eigenvalue as we want to minimize the Euclidean distance above.

Q8-

From the question: $\mathrm{p}(\omega_1)\Sigma_1 + \mathrm{p}(\omega_1)m_1 m_1^T + \mathrm{p}(\omega_2)\Sigma_2 + \mathrm{p}(\omega_2)m_2 m_2^T = I$ $\mathrm{p}(\omega_1)\mathrm{m}_1 + \mathrm{p}(\omega_2)m_2 = 0$

Part A) $\mathrm{p}(\omega_1)\Sigma_1 + \mathrm{p}(\omega_2)\Sigma_2 = I - \mathrm{p}(\omega_1)m_1 m_1^T - \mathrm{p}(\omega_2)m_2 m_2^T = I - \mathrm{p}(\omega_1)m_1 m_1^T - \mathrm{p}(\omega_2)(\frac{-\mathrm{p}(\omega_1)}{\mathrm{p}(\omega_2)})m_1(\frac{-\mathrm{p}(\omega_1)}{\mathrm{p}(\omega_2)})m_1^T$

$[\mathrm{p}(\omega_1)\Sigma_1 + \mathrm{p}(\omega_2)\Sigma_2]^{-1} = [I - \mathrm{p}(\omega_1)m_1 m_1^T - \frac{\mathrm{p}(\omega_1)^2}{\mathrm{p}(\omega_2)}m_1 m_1^T]^{-1} = [I - \frac{\mathrm{p}(\omega_1)\mathrm{p}(\omega_2)+\mathrm{p}(\omega_1)^2}{\mathrm{p}(\omega_2)}m_1 m_1^T]^{-1}$

Part B) $(m_2 - m_1) = \frac{-\mathrm{p}(\omega_1)}{\mathrm{p}(\omega_2)}m_1 - m_1 = -m_1\left(\frac{\mathrm{p}(\omega_1)+\mathrm{p}(\omega_2)}{\mathrm{p}(\omega_2)}\right) = -m_1\left(\frac{1}{\mathrm{p}(\omega_2)}\right) = \frac{-m_1}{\mathrm{p}(\omega_2)}$

$(m_2 - m_1)^T[\mathrm{p}(\omega_1)\Sigma_1 + \mathrm{p}(\omega_2)\Sigma_2]^{-1}(m_2 - m_1) = \frac{-m_1{}^T}{\mathrm{p}(\omega_2)}[I - \frac{\mathrm{p}(\omega_1)\mathrm{p}(\omega_2)+\mathrm{p}(\omega_1)^2}{\mathrm{p}(\omega_2)}m_1 m_1^T]^{-1}\frac{-m_1}{\mathrm{p}(\omega_2)}$ $m_1 = \begin{bmatrix} m_{11} \\ m_{12} \end{bmatrix}$

Part C) $|\mathrm{p}(\omega_1)\Sigma_1 + \mathrm{p}(\omega_2)\Sigma_2| = \left|I - \frac{\mathrm{p}(\omega_1)\mathrm{p}(\omega_2)+\mathrm{p}(\omega_1)^2}{\mathrm{p}(\omega_2)}m_1 m_1^T\right| = |I - A m_1 m_1^T| = \begin{vmatrix} 1 - Am_{11}^2 & Am_{11}m_{12} \\ Am_{11}m_{12} & 1 - Am_{12}^2 \end{vmatrix} = 1 - A(m_{11}^2 + m_{12}^2) + A^2 m_{11}^2 m_{12}^2 - A^2 m_{11}^2 m_{12}^2 = 1 - A(m_{11}^2 + m_{12}^2) = 1 - A\|m_1\|^2$

Q10-

In order to simultaneously diagonalization of Σ_1 and Σ_2, the simplest way (in my opinion!) is finding the eigenvector matrix of $\Sigma_1^{-1}\Sigma_2$:

$$\Sigma_1^{-1}\Sigma_2 = \frac{1}{1-\frac{1}{4}}\begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}\begin{bmatrix} 1+\frac{\sqrt{3}}{4} & 0.5 \\ 0.5 & 1-\frac{\sqrt{3}}{4} \end{bmatrix} = \frac{4}{3}\begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}\begin{bmatrix} 1+\frac{\sqrt{3}}{4} & 0.5 \\ 0.5 & 1-\frac{\sqrt{3}}{4} \end{bmatrix} = \begin{bmatrix} \frac{4}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{4}{3} \end{bmatrix}\begin{bmatrix} 1+\frac{\sqrt{3}}{4} & 0.5 \\ 0.5 & 1-\frac{\sqrt{3}}{4} \end{bmatrix} =$$

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

Calculating eigenvalue and eigenvector matrices from $\Phi(\Sigma - \Lambda) = 0$. We will have:

$$\Lambda = \begin{bmatrix} 1.5 & 0 \\ 0 & 0.5 \end{bmatrix} \quad \text{and} \quad \Phi = \begin{bmatrix} 0.9659 & -0.2588 \\ -0.2588 & 0.9659 \end{bmatrix}$$

Diagonalization:

$$\Phi^T\Sigma_1\Phi = \begin{bmatrix} 0.9659 & -0.2588 \\ -0.2588 & 0.9659 \end{bmatrix}\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\begin{bmatrix} 0.9659 & -0.2588 \\ -0.2588 & 0.9659 \end{bmatrix} = \begin{bmatrix} 0.75 & 0 \\ 0 & 0.75 \end{bmatrix}$$

$$\Phi^T\Sigma_2\Phi = \begin{bmatrix} 0.9659 & -0.2588 \\ -0.2588 & 0.9659 \end{bmatrix}\begin{bmatrix} 1+\frac{\sqrt{3}}{4} & 0.5 \\ 0.5 & 1-\frac{\sqrt{3}}{4} \end{bmatrix}\begin{bmatrix} 0.9659 & -0.2588 \\ -0.2588 & 0.9659 \end{bmatrix} = \begin{bmatrix} 1.125 & 0 \\ 0 & 0.375 \end{bmatrix}$$

Q11-

$\Sigma = S - mm^T$ $\qquad$ $(A-B)^{-1} = A^{-1} - \frac{1}{1+trace(BA^{-1})}A^{-1}BA^{-1}$ $\qquad$ $trace(AB) = trace(BA)$

$|X + AB| = |X||I + BX^{-1}A|$ $\qquad$ $|I + AB| = |I + BA|$

$|I + cr| = 1 + rc$ $\qquad$ Where c is a column matrix and r a row matrix.

From the useful formulae above: $\qquad$ $\Sigma^{-1} = (S - mm^T)^{-1} = S^{-1} - \frac{1}{1+trace(mm^TS^{-1})}S^{-1}mm^TS^{-1}$

$$\Sigma^{-1}m = S^{-1}m - \frac{1}{1+trace(mm^TS^{-1})}S^{-1}mm^TS^{-1}m = S^{-1}m(1 - \frac{1}{1+trace(mm^TS^{-1})}m^TS^{-1}m)$$

$$trace(mm^TS^{-1}) = trace(m^TS^{-1}m) = m^TS^{-1}m$$

$$\Sigma^{-1}m = S^{-1}m\left(1 - \frac{1}{1+m^TS^{-1}m}m^TS^{-1}m\right) = S^{-1}m(\frac{1}{1+m^TS^{-1}m})$$

On the other hand: $|\Sigma| = |S - mm^T| = |S||I + m^TS^{-1}m| = |S||I + S^{-1}mm^T| = |S|(1 + m^TS^{-1}m)$

So, $\Sigma^{-1}m = S^{-1}m\left(\frac{1}{1+m^TS^{-1}m}\right) = S^{-1}m\frac{|S|}{|\Sigma|}$ $\qquad$ Therefore: $\Sigma^{-1}m = \frac{|S|}{|\Sigma|}S^{-1}m$

Q12-

$\sum \lambda_i = \lambda = trace(\Sigma^{-1} m m^T) = trace(m^T \Sigma^{-1} m) = m^T \Sigma^{-1} m$

$\Sigma^{-1} m m^T \phi = \phi\lambda = \phi m^T \Sigma^{-1} m$ The eigenvector that makes the equation true is $\phi = \Sigma^{-1} m$

Q13-

Part A. $J(\varphi) = \varphi^T \Sigma \varphi - \lambda(\varphi^T \varphi - 1)$ $\frac{\partial J(\varphi)}{\partial \varphi} = 2\Sigma\varphi - 2\lambda\varphi = 0$ $\Sigma\varphi = \lambda\varphi$

φ is the eigenvector of matrix Σ and it is corresponding to the maximum eigenvalue since we want to minimize the cost function.

Part B.

$\varphi_{k+1} = \varphi_k - \mu(2\Sigma\varphi - 2\lambda\varphi) = \varphi_k - \mu(2\Sigma\varphi - 2\varphi^T \Sigma\varphi\varphi) = \varphi_k - \mu(\Sigma\varphi - (\varphi^T \Sigma\varphi)\varphi)$

What happened to 2? I defined a new variable: 2μ and named it μ again!

Part C.

$\boldsymbol{\Sigma}_0 = \boldsymbol{\varepsilon}\mathbf{I}$

$\boldsymbol{\varphi}_0 = \boldsymbol{\alpha}[1\ 1 \ldots 1]^t$

$\hat{\mathbf{m}}_0 = [0\ 0 \ldots 0]^t$

$\mu = \mu_0$

for k=1:N

$\hat{\mathbf{m}}_k = \frac{(k-1)\,\hat{\mathbf{m}}_{k-1} + \mathbf{x}_k}{k}$

$\hat{\boldsymbol{\Sigma}}_k = (1 - (k-1)/k)(\mathbf{x}_k - \hat{\mathbf{m}}_k)(\mathbf{x}_k - \hat{\mathbf{m}}_k)^t + ((k-1)/k)\,\hat{\boldsymbol{\Sigma}}_{k-1}$

$\boldsymbol{\varphi}_k = \boldsymbol{\varphi}_{k-1} - \mu\left(\hat{\boldsymbol{\Sigma}}_k \boldsymbol{\varphi}_{k-1} - \left(\boldsymbol{\varphi}_{k-1}^t \hat{\boldsymbol{\Sigma}}_k \boldsymbol{\varphi}_{k-1}\right)\boldsymbol{\varphi}_{k-1}\right)$

end

Q14-

$$\Sigma = \Phi\Lambda\Phi^{-1} \rightarrow \Sigma^{-1} = (\Phi\Lambda\Phi^{-1})^{-1} = ((\Phi\Lambda)(\Phi^{-1}))^{-1} = \Phi(\Phi\Lambda)^{-1} = \Phi\Lambda^{-1}\Phi^{-1}$$

Q15-

$D_A = P^{-1}AP \rightarrow A = PD_AP^{-1}$

$D_B = P^{-1}BP \rightarrow B = PD_BP^{-1}$

$AB = PD_AP^{-1}PD_BP^{-1} = PD_AD_BP^{-1} = PD_BD_AP^{-1} = PD_BP^{-1}PD_AP^{-1} = PD_BP^{-1}PD_AP^{-1} = BA$

Q16-

The covariance matrix Σ has orthonormal basis $\phi_1, \phi_2, \ldots, \phi_n$.

A data point can be written as $v = c_1\phi_1 + c_2\phi_2 + \cdots + c_n\phi_n$

$$\Sigma^m v = c_1\Sigma^m\phi_1 + c_2\Sigma^m\phi_2 + \cdots + c_n\Sigma^m\phi_n$$

Remember $\Sigma\phi_i = \lambda_i\phi_i$, $\Sigma^m\phi_i = \lambda_i{}^m\phi_i$

$$\Sigma^m v = c_1\lambda_1{}^m\phi_1 + c_2\lambda_2{}^m\phi_2 + \cdots + c_n\lambda_n{}^m\phi_n$$

Lets find the angle between $\Sigma^m v$ and ϕ_1:

$$\cos\theta = \frac{\Sigma^m v \,.\, \phi_1}{\|\Sigma^m v\|} = \frac{(\Sigma^m v)^t\phi_1}{\|\Sigma^m v\|} = \frac{(c_1\lambda_1{}^m\phi_1 + c_2\lambda_2{}^m\phi_2 + \cdots + c_n\lambda_n{}^m\phi_n)^t\phi_1}{\|\Sigma^m v\|}$$
$$= \frac{c_1\lambda_1{}^m\phi_1{}^t\phi_1 + c_2\lambda_2{}^m\phi_2{}^t\phi_1 + \cdots + c_n\lambda_n{}^m\phi_n{}^t\phi_1}{\|\Sigma^m v\|} = \frac{c_1\lambda_1{}^m}{\sqrt{(\Sigma^m v)^t(\Sigma^m v)}}$$
$$= \frac{c_1\lambda_1{}^m}{\sqrt{c_1{}^2\lambda_1{}^{2m} + c_2{}^2\lambda_2{}^{2m} + \cdots + c_n{}^2\lambda_n{}^{2m}}}$$

With increasing m, $c_1\lambda_1{}^m$ gets larger and we can ignore the other terms. This leads to $\cos\theta = 1 \rightarrow \theta = 0$