Answer Key and Marking Scheme

Q1

(A) i) The line is $x-2y+3=0$. The optimization problem to be minimized to obtain the closest point is given by

$$\min\,(4-x)^2+(2-y)^2$$
$$\text{subject to } x-2y+3=0$$

(0.5 marks)

Construct the lagrangian equation as

$$L(x,y,\lambda)=(4-x)^2+(2-y)^2+\lambda(x-2y+3)=0$$

Solve the equation $\nabla L(x,y,\lambda)=\mathbf{0}$
$2(x-4)+\lambda=0, 2(y-2)-2\lambda=0, x-2y+3=0$ (0.5 marks)

Then , $x=\frac{17}{5}, y=\frac{16}{5}$ (0.5 marks)

And distance $=\sqrt{(4-\frac{17}{5})^2+(2-\frac{16}{5})^2}=\sqrt{\frac{9}{5}}$ (0.5 marks)

ii) The line is $x+2y+5=0$. The optimization problem to be minimized to obtain the closest point is given by

$$\min\,(4-x)^2+(2-y)^2$$
$$\text{subject to } x+2y+5=0$$

(0.5 marks)

Construct the lagrangian equation as

$$L(x,y,\lambda)=(4-x)^2+(2-y)^2+\lambda(x+2y+5)=0$$

Solve the equation $\nabla L(x,y,\lambda)=\mathbf{0}$
$2(x-4)+\lambda=0, 2(y-2)+2\lambda=0, x+2y+5=0$ (0.5 marks)

Then , $x=\frac{7}{5}, y=\frac{-16}{5}$ (0.5 marks)

And distance $=\sqrt{(4-\frac{7}{5})^2+(2-\frac{-16}{5})^2}=\sqrt{\frac{169}{5}}$ (0.5 marks)

(B) For the given matrix C , we will first find the eigenvalues, so that we can use the trace and determinant property :
Construct equation $det(C-\lambda I)=0$. The resultant characteristic equation is given by

$$det(C-\lambda I)=\lambda^3-14\lambda^2-9\lambda+126$$

(0.5 mark)

Factorize to get $det(C-\lambda I)=(\lambda-3)(\lambda+3)(\lambda-14)=0$. The eigenvalues of C is given by: $\lambda_1=14, \lambda_2=3, \lambda_3=-3$ (0.5 mark)

i) Recall that if λ is an eigenvalue of C, then λ^n is an eigenvalue of C^n.

Hence, λ_1^6, λ_2^6, λ_3^6 are the 3 eigenvalues of C^6. (0.5 mark)
Hence

$$
\begin{aligned}
Trace(C^6) &= \lambda_1^6 + \lambda_2^6 + \lambda_3^6 \\
&= (14)^6 + (3)^6 + (-3)^6
\end{aligned}
$$

(0.5 mark)

ii) Recall that $det(AB) = det(A)det(B)$.
Hence $det(C^7) = det(C)^7$. (0.5 mark)

We know that det(C)= $\lambda_1 * \lambda_2 * \lambda_3 = 14 * 3 * (-3) = -126$.

$$
det(C^7) = det(C)^7 = (-126)^7
$$

(0.5 mark)

(C) Given that $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are linearly independent. Hence $\beta_1\mathbf{a} + \beta_2\mathbf{b} + \beta 3\mathbf{c} = \mathbf{0}$ will have only trivial solution $\beta_1 = \beta_2 = \beta_3 = 0$. (0.5 marks)
Now consider $\alpha_1\mathbf{x} + \alpha_2\mathbf{y} + \alpha_3\mathbf{z} = 0$.

$$
\begin{aligned}
\alpha_1\mathbf{x} + \alpha_2\mathbf{y} + \alpha_3\mathbf{z} = \alpha_1(\mathbf{b} - \mathbf{c}) + \alpha_2(\mathbf{a} + \mathbf{c}) + \alpha_3(\mathbf{a} - \mathbf{b}) = 0 \\
\mathbf{a}(\alpha_2 + \alpha_3) + \mathbf{b}(\alpha_1 - \alpha_3) + \mathbf{c}(\alpha_2 - \alpha_1) = 0
\end{aligned}
$$

(1 marks)

Since $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are linearly independent, we can conclude that
$\alpha_2 + \alpha_3 = 0$,
$\alpha_1 - \alpha_3 = 0$, ,
$\alpha_2 - \alpha_1 = 0$ (0.5 marks)
Solving for $\alpha_1, \alpha_2, \alpha_3$ from the 3 equations, we get
$\alpha_1 = 0, \alpha_2 = 0, \alpha_3 = 0$ (0.5 marks)

Hence $\{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$ is a linearly independent set. (0.5 marks)

Q2

(A) i) We have

$$\begin{aligned} L(\boldsymbol{\beta}) &= \frac{1}{2p}\|\boldsymbol{y}-\boldsymbol{\beta}\|^2+\lambda\|\boldsymbol{W}\boldsymbol{\beta}\|^2 \\ &= \frac{1}{p}(\sum_{j=1}^{p}[\frac{1}{2}(y_j-\beta_j)^2+\lambda\|\boldsymbol{W}\boldsymbol{\beta}\|^2]) \\ &= \frac{1}{p}\sum_{j=1}^{p}L_j(\boldsymbol{\beta}) \text{ where} \\ & L_j(\boldsymbol{\beta})=\frac{1}{2}(y_j-\beta_j)^2+\lambda\|\boldsymbol{W}\boldsymbol{\beta}\|^2 \end{aligned}$$

(1.5 marks)

ii) Now $L_j(\boldsymbol{\beta})=\frac{1}{2}(y_j-\beta_j)^2+\lambda\|\boldsymbol{W}\boldsymbol{\beta}\|^2=(y_j-\beta_j)^2+\lambda\boldsymbol{\beta}^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{\beta}$.
We have

$$\frac{\partial[\frac{1}{2}(y_j-\beta_j)^2]}{\partial\beta_i}=v_i$$

where

$$v_i=\begin{cases} 0 & \text{when } i\neq j \\ -(y_j-\beta_j) & \text{when } i=j \end{cases}$$

(1 mark)

Also

$$\frac{\partial\lambda\boldsymbol{\beta}^T\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{\beta}}{\partial\boldsymbol{\beta}}=2\lambda\boldsymbol{\beta}^T\boldsymbol{W}^T\boldsymbol{W}=(2\lambda\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{\beta})^T.$$

(0.5 marks)

Thus, we have

$$\nabla L_j(\boldsymbol{\beta})=\left(\boldsymbol{v}+2\lambda\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{\beta}\right)^T, j=1,\cdots p.$$

where $\boldsymbol{v}=[v_1,\cdots,v_p]^T$ is a p dimensional vector such that

$$v_i=\begin{cases} 0 & \text{when } i\neq j \\ -(y_j-\beta_j) & \text{when } i=j \end{cases}$$

(0.5 marks)

iii) Now $L(\boldsymbol{\beta})=\frac{1}{p}\sum_{j=1}^{p}L_j(\boldsymbol{\beta})$. Therefore,

$$\begin{aligned} \nabla L(\boldsymbol{\beta}) &= \frac{1}{p}\sum_{j=1}^{p}\nabla L_j(\boldsymbol{\beta}) \\ &= -\frac{1}{p}(\boldsymbol{y}-\boldsymbol{\beta})^T+2\lambda(\boldsymbol{W}^T\boldsymbol{W}\boldsymbol{\beta})^T. \end{aligned}$$

(1.5 marks)

(B) The dimension of the transformed data $=1+m+m+m-2=3m-1$.
(0.5 marks)

$$
\begin{aligned}
\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{y}) &= \boldsymbol{\phi}(\boldsymbol{x})^T \boldsymbol{\phi}(\boldsymbol{y}) \\
&= 1 + x_1 y_1 + \cdots + x_m y_m + x_1^3 y_1^3 + \cdots + x_m^3 y_m^3 + x_1 x_2 x_3 y_1 y_2 y_3 \\
&\quad + x_2 x_3 x_4 y_2 y_3 y_4 + \cdots + x_{m-2} x_{m-1} x_m y_{m-2} y_{m-1} y_m.
\end{aligned}
$$

(1 mark)

(C) We have $f(x, y) = (x^2 + y^2)e^{-(x^2+y^2)}$ and A is the set of all possible critical points of f. So,

i) To find critical points,

$$
\begin{aligned}
f_x(x, y) &= 2xe^{-(x^2+y^2)}[1 - x^2 - y^2] = 0 \\
\Rightarrow &\quad 2x[1 - x^2 - y^2] = 0 \\
\Rightarrow &\quad x = 0 \text{ or } x^2 + y^2 = 1. \text{ (0.5 marks)} \\
f_x(x, y) &= 2ye^{-(x^2+y^2)}[1 - x^2 - y^2] = 0 \\
\Rightarrow &\quad 2y[1 - x^2 - y^2] = 0 \\
\Rightarrow &\quad y = 0 \text{ or } x^2 + y^2 = 1. \text{ (0.5 marks)}
\end{aligned}
$$

Therefore, $A = \{(0, 0)\} \cup \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 = 1\}$. (0.5 marks)

ii) To find the nature of $(0, 0)$, consider

$$
\begin{aligned}
f_{xx}(x, y) &= 2e^{-(x^2+y^2)} - 4x^2 e^{-(x^2+y^2)} - (6x^2 + 2y^2)e^{-(x^2+y^2)} + 4x^2(x^2 + y^2)e^{-(x^2+y^2)} \\
\Rightarrow f_{xx}(0, 0) &= 2. \text{ (0.5 marks)} \\
f_{yy}(x, y) &= 2e^{-(x^2+y^2)} - 4y^2 e^{-(x^2+y^2)} - (2x^2 + 6y^2)e^{-(x^2+y^2)} + 4y^2(x^2 + y^2)e^{-(x^2+y^2)} \\
\Rightarrow f_{yy}(0, 0) &= 2. \text{ (0.5 marks)} \\
f_{xy}(x, y) &= f_{yx}(x, y) \\
&= -8xye^{-(x^2+y^2)} + 4xy(x^2 + y^2)e^{-(x^2+y^2)} \\
\Rightarrow f_{xy}(0, 0) &= f_{yx}(0, 0) = 0. \text{ (0.5 marks)}
\end{aligned}
$$

Clearly $H(0, 0) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$ and hence positive definite as its eigenvalues are $2, 2$. Therefore, the $(0, 0)$ is a point of minima. (0.5marks)

Q3

(A) i) The claim is not true. (1.5 marks for finding a matrix satisfying all conditions.
1.5 marks for finding RREF of that matrix with $\boldsymbol{C_1}$ as one of the pivotal columns.)

ii) Now

$$\begin{aligned}\boldsymbol{C_1} &= \sum_{i=2}^{6} \boldsymbol{C_i} \\ &= \boldsymbol{C_2} + \sum_{i=3}^{6} \boldsymbol{C_i} \\ &= 2\boldsymbol{C_2}.\end{aligned}$$

(0.5 marks)

$\Rightarrow \boldsymbol{C_1} - 2\boldsymbol{C_2} + 0\boldsymbol{C_3} + 0\boldsymbol{C_4} + 0\boldsymbol{C_5} + 0\boldsymbol{C_6} = \boldsymbol{0}$
$\Rightarrow [\boldsymbol{C_1}, \boldsymbol{C_2}, \boldsymbol{C_3}, \boldsymbol{C_4}, \boldsymbol{C_5}, \boldsymbol{C_6}][1, -2, 0, 0, 0, 0]^T = \boldsymbol{0}$
$\Rightarrow \boldsymbol{A}[1, -2, 0, 0, 0, 0]^T = \boldsymbol{0}.$ (0.5 marks)
Similarly

$$\begin{aligned}\boldsymbol{C_1} &= \sum_{i=2}^{6} \boldsymbol{C_i} \\ &= \boldsymbol{C_2} + \sum_{i=3}^{6} \boldsymbol{C_i} \\ &= 2\sum_{i=3}^{6} \boldsymbol{C_i}.\end{aligned}$$

(0.5 marks)

$\Rightarrow -1\boldsymbol{C_1} + 0\boldsymbol{C_2} + 2\boldsymbol{C_3} + 2\boldsymbol{C_4} + 2\boldsymbol{C_5} + 2\boldsymbol{C_6} = \boldsymbol{0}$
$\Rightarrow [\boldsymbol{C_1}, \boldsymbol{C_2}, \boldsymbol{C_3}, \boldsymbol{C_4}, \boldsymbol{C_5}, \boldsymbol{C_6}][-1, 0, 2, 2, 2, 2]^T = \boldsymbol{0}$
$\Rightarrow \boldsymbol{A}[-1, 0, 2, 2, 2, 2]^T = \boldsymbol{0}.$ (0.5 marks)

(B) i) The problem is not linearly separable as its difficult to find the linear decision boundary which separates the two classes. (1 mark)

ii) The Kernel is defined as: $K(x_1, x_2) = (4 + x_1^T x_2)^2$ (0.5 mark)
The Kernel matrix is defined as

$$K = \begin{pmatrix} 144 & 16 & 16 & 16 \\ 16 & 144 & 16 & 16 \\ 16 & 16 & 144 & 16 \\ 16 & 16 & 16 & 144 \end{pmatrix}$$

(1 mark)

iii)

$$Q(\alpha) = \sum_{i=1}^{4} \alpha_i - \frac{1}{2}(\sum_{i=1}^{4} 144\alpha_i^2 + 32\alpha_1\alpha_2 + 32\alpha_3\alpha_4 - 32\alpha_1\alpha_3 - 32\alpha_1\alpha_4 - 32\alpha_2\alpha_3 - 32\alpha_2\alpha_4)$$

$$\frac{\partial Q(\alpha)}{\partial \alpha_i} = 0$$

We get system of equations:

$$\begin{aligned}
144\alpha_1 + 16\alpha_2 - 16\alpha_3 - 16\alpha_4 &= 1 \\
16\alpha_1 + 144\alpha_2 - 16\alpha_3 - 16\alpha_4 &= 1 \\
-16\alpha_1 - 16\alpha_2 + 144\alpha_3 + 16\alpha_4 &= 1 \\
-16\alpha_1 - 16\alpha_2 + 16\alpha_3 + 144\alpha_4 &= 1
\end{aligned}$$

solving above equations we get $\alpha_i = \frac{1}{128}$ (0.5 marks)
$w = (0, 0, 22.6, 0, 0, 0)$ and $b = 0$
The decision boundary is given by

$$w^T \phi(x) + b = 0$$

$$x_1 x_2 = 0 \qquad \text{(2 marks)}$$

Q4

(A) i) The Lagrangian is given by $L(x,\lambda) = x_1^2 + x_2^2 - 4x_1 - 4x_2 + \lambda_1(x_1^2 - x_2) + \lambda_2(x_1 + x_2 - 2)$
$\frac{\partial L}{\partial x_1} = 0$, $\frac{\partial L}{\partial x_2} = 0$

$x_1 = \frac{(4-\lambda_2)}{2(1+\lambda_1)}$ and $x_2 = \frac{\lambda_1-\lambda_2+4}{2}$ (1 mark)

Substituting above x_1 and x_2 values in $L(x,\lambda)$
$L(\lambda) = \frac{(4-\lambda_2)^2}{4(1+\lambda_1)^2} + \frac{(\lambda_1-\lambda_2+4)^2}{4} - 4\frac{(4-\lambda_2)}{2(1+\lambda_1)} - 4\frac{\lambda_1-\lambda_2+4}{2} + \lambda_1(\frac{(4-\lambda_2)^2}{4(1+\lambda_1)^2} - \frac{\lambda_1-\lambda_2+4}{2}) + \lambda_2(\frac{(4-\lambda_2)}{2(1+\lambda_1)} + \frac{\lambda_1-\lambda_2+4}{2} - 2)$ (1 mark)

Dual is given by: $\max_\lambda L(\lambda)$

ii) Dual and primal objective function value will be the same as the primal objective function is convex and the constraints are convex. (2 marks)

(B) i) the dimension of each training sample is 1×6 as the covariance matrix is having 6 eigen values so that means each training sample will have 6 columns. (1 mark)

ii) total variance= sum of all eigen values=23.330 if we want to retain 95% variance then we should keep top 3 eigen values as (12 +6.8 +3.5)*100/(23.330)=95.7 (1.5 marks)

if we want to retain 99% variance then we should keep top 4 eigen values as (12 +6.8 +3.5+1)*100/(23.330)=99.87 (1.5 marks)

iii) After PCA the result will be $\begin{bmatrix}-3 & -2 & 2 & 1 & 2 & 4\end{bmatrix} * \begin{pmatrix}0.115 & 0.205\\ 0.106 & 0.215\\ 0.118 & 0.315\\ 0.082 & 0.428\\ 0.013 & 0.238\\ 0.023 & 0.034\end{pmatrix} =$

$\begin{bmatrix}-0.1210 & 0.6250\end{bmatrix}$ (2 marks)