Q Answer the following questions with justifications.

(A) Suppose $\boldsymbol{A} = [\boldsymbol{C}_1, \boldsymbol{C}_2, \boldsymbol{C}_3, \boldsymbol{C}_4]$ where $\boldsymbol{C}_m \in \mathbb{R}^4, m = 1, 2, 3, 4$ are columns of $\boldsymbol{A}$. It is known that $\text{rank}(\boldsymbol{A}) = 2$ and $\boldsymbol{C}_2 = 3\boldsymbol{C}_1$ and $\boldsymbol{C}_4 = 2\boldsymbol{C}_1 + 3\boldsymbol{C}_3$. If a particular solution of $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ is $[1, 0, 1, 0]^T$, then

(i) Find the general solution of $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$. [2 Marks]

(ii) Find $\boldsymbol{b}$ if $\text{RREF}(\boldsymbol{A}) = \boldsymbol{A}$. [2 Marks]

(B) Consider $M = \{\boldsymbol{A} \in \mathbb{R}^{2\times 2} \mid \boldsymbol{A} = -\boldsymbol{A}^T\}$.

(i) Prove that M is subspace of vector space V, where V is set of all 2×2 real matrices. [1.5 Marks]

(ii) Prove or disprove $\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$ is a basis for M defined in (i). [2.5 Marks]

(iii) Prove or disprove that $\left\{ \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 5 & -5 \\ 5 & 5 \end{bmatrix}, \begin{bmatrix} 0 & -3 \\ 3 & 1 \end{bmatrix} \right\}$ is a linearly independent set. [2 Marks]

Solution

(A) Now $\boldsymbol{A} = [\boldsymbol{C}_1, \boldsymbol{C}_2, \boldsymbol{C}_3, \boldsymbol{C}_4]$ where $\boldsymbol{C}_m \in \mathbb{R}^4, m = 1, 2, 3, 4$ are columns of $\boldsymbol{A}$. It is known that $\text{rank}(\boldsymbol{A}) = 2$

(i) $\boldsymbol{C}_2 = 3\boldsymbol{C}_1 \Rightarrow 3\boldsymbol{C}_1 - \boldsymbol{C}_2 + 0\boldsymbol{C}_3 + 0\boldsymbol{C}_4 = \boldsymbol{0}$

$$\Rightarrow [\boldsymbol{C}_1, \boldsymbol{C}_2, \boldsymbol{C}_3, \boldsymbol{C}_4] \begin{bmatrix} 3 \\ -1 \\ 0 \\ 0 \end{bmatrix}$$ [0.5 Marks]

$\boldsymbol{C}_4 = 2\boldsymbol{C}_1 + 3\boldsymbol{C}_3 \Rightarrow 2\boldsymbol{C}_1 + 0\boldsymbol{C}_2 + 3\boldsymbol{C}_3 - 1\boldsymbol{C}_4 = \boldsymbol{0}$

$$\Rightarrow [\boldsymbol{C}_1, \boldsymbol{C}_2, \boldsymbol{C}_3, \boldsymbol{C}_4] \begin{bmatrix} 2 \\ 0 \\ 3 \\ -1 \end{bmatrix}$$ [0.5 Marks]

Thus, the general solution of $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ is given by

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} + \lambda_1 \begin{bmatrix} 3 \\ -1 \\ 0 \\ 0 \end{bmatrix} + \lambda_2 \begin{bmatrix} 2 \\ 0 \\ 3 \\ -1 \end{bmatrix} \mid \forall \lambda_1, \lambda_2 \in \mathbb{R} \right\}$$ [0.5 Marks]

(ii) Now $\boldsymbol{A} = [\boldsymbol{C}_1, \boldsymbol{C}_2, \boldsymbol{C}_3, \boldsymbol{C}_4] = [\boldsymbol{C}_1, 2\boldsymbol{C}_1, \boldsymbol{C}_3, 2\boldsymbol{C}_1 + 3\boldsymbol{C}_3]$ and $\text{rank}(\boldsymbol{A}) = 2$.
Therefore $\boldsymbol{C}_1, \boldsymbol{C}_3$ are pivot columns and $\text{RREF}(\boldsymbol{A}) = \boldsymbol{A}$
$\Rightarrow \boldsymbol{C}_1 = [1, 0, 0, 0]^T$ and $\boldsymbol{C}_3 = [0, 1, 0, 0]^T$
$\Rightarrow \boldsymbol{C}_2 = [3, 0, 0, 0]^T$ and $\boldsymbol{C}_4 = [2, 3, 0, 0]^T$

$$\Rightarrow \boldsymbol{A} = \begin{bmatrix} 1 & 3 & 0 & 2 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$ [1 Mark]

Therefore $\boldsymbol{b} = \boldsymbol{A}[1,0,1,0]^T = [1,1,0,0]^T$ [0.5 Marks]

(B) (i) Clearly the zero matrix of order 2, $\mathbf{0} \in M \Rightarrow M \neq \phi$. [0.5 Marks]

Let $\boldsymbol{A}, \boldsymbol{B} \in M, k \in \mathbb{R}$

$\Rightarrow \boldsymbol{A} = -\boldsymbol{A}^T, \boldsymbol{B} = -\boldsymbol{B}^T$

$\Rightarrow \boldsymbol{A} + \boldsymbol{B} = -\boldsymbol{A}^T - \boldsymbol{B}^T = -(\boldsymbol{A} + \boldsymbol{B})^T$

$\Rightarrow \boldsymbol{A} + \boldsymbol{B} \in M$. [0.5 Marks]

$k\boldsymbol{A} = k(-\boldsymbol{A}^T) = (-k)\boldsymbol{A}^T = -(k\boldsymbol{A})^T$

So, M is a subspace. [0.5 Marks]

(ii) Clearly $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$ does not belong to M and hence cannot be a basis of M.

(Kindly award full marks for any valid reason for disproving the claim.) [2.5 Marks]

(iii) Consider $a \begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix} + b \begin{bmatrix} 5 & -5 \\ 5 & 5 \end{bmatrix} + c \begin{bmatrix} 0 & -3 \\ 3 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$ [0.5 Marks]

$$\begin{aligned} \Rightarrow a + 5b &= 0 \\ -a - 5b - 3c &= 0 \\ a + 5b + 3c &= 0 \\ 5b + c &= 0 \end{aligned}$$

$\Rightarrow a = b = c = 0$. Hence it a linearly independent set. [1.5 Marks]

(Kindly award full marks for any other correct method.)

Q Answer the following questions with justifications.

(A) Consider the following optimization problem (A) on the data $(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots (\boldsymbol{x}_n, y_n)$ of the following form:

$$\max \sum_{i=1}^{i=n} \alpha_i - \sum_{i=1}^{i=n} \sum_{j=1}^{j=n} \alpha_i \alpha_j \boldsymbol{x}_i . \boldsymbol{x}_j$$

$$\text{subject to} \sum_{i=1}^{i=n} \alpha_i y_i = 0$$

$$\alpha_i \geq 0, \forall i$$

Note that here y_i is $\pm 1, \forall i$ and each $\boldsymbol{x}_i$ is a $n \times 1$ vector. The variables in the problem are the α_i. We form a new optimization problem (B) as follows:

$$\max \sum_{i=1}^{i=n+1} \alpha_i - \sum_{i=1}^{i=n+1} \sum_{j=1}^{j=n+1} \alpha_i \alpha_j \boldsymbol{x}_i . \boldsymbol{x}_j$$

$$\text{subject to} \sum_{i=1}^{i=n+1} \alpha_i y_i = 0$$

$$\alpha_i \geq 0, \forall i, 1 \leq i \leq n+1$$

where $\boldsymbol{x}_{n+1} = \frac{1}{2}(\boldsymbol{x}_i + \boldsymbol{x}_j)$ for some values of i and j such that $y_i = y_j$. We set $y_{n+1} = y_i$. Show that the maximum value of the objective function for the problem (B) is greater than or equal to the maximum value of the objective function for problem (A). Justify your solution mathematically. [3 Marks]

(B) Let $\boldsymbol{x}$ and $\boldsymbol{z}$ be two $n \times 1$ vectors, for which a kernel function is defined as $K(\boldsymbol{x}, \boldsymbol{z}) = (\boldsymbol{x}^T \boldsymbol{z})^2 + 3(\boldsymbol{x}^T \boldsymbol{z} + 2)^2$. If possible, find a mapping ϕ from the space of $n \times 1$ vectors to the space of $(n^2 + n + 1) \times 1$ vectors for which the given Kernel function represents the inner product. Otherwise explain why such a mapping is not possible for the given kernel function. [5 Marks]

(C) Consider a gradient update rule given by:

$$a_{t+1} = \gamma a_t + (1 - \gamma) \nabla_w (L)$$

$$w_{t+1} = w_t - a_{t+1}$$

What is the contribution of a_0 while computing the value of a_5 [2 Marks]

Solutions

(A) Let $(\alpha_1^*, \alpha_2^*, \ldots \alpha_n^*)$ be the optimal solution for problem (A), and let the optimal objective value be O_A. For problem (B) let us set $(\alpha_1, \alpha_2, \ldots \alpha_n, \alpha_{n+1}) = (\alpha_1^*, \alpha_2^*, \ldots \alpha_n^*, 0)$. It is easy to see that this tuple is a feasible solution for problem (B) since both the constraints are satisfied. The value of the objective function of problem (B) for this assignment to the α_is can be seen to equal to O_A, since the terms in the objective function of (B) that do not exist in the objective function for (A) contain α_{n+1} which is set to zero in the feasible solution for (B). Thus the optimal solution problem has to greater than or equal to O_A, since the optimal solution is greater than or equal to any feasible solution to the problem.

(B) The given kernel function can be written as $(\boldsymbol{x}^T\boldsymbol{z})^2 + 3(\boldsymbol{x}^T\boldsymbol{z} + 2)^2 = 4(\boldsymbol{x}^T\boldsymbol{z})^2 + 12\boldsymbol{x}^T\boldsymbol{z} + 12$. This can be seen as an inner product ϕ of the following form $\phi = [\phi_1(\boldsymbol{x}), \phi_2(\boldsymbol{x}), \phi_3(\boldsymbol{x})]^T$ where ϕ_1 is a $n^2 \times 1$ mapping representing the term $4(\boldsymbol{x}^T\boldsymbol{z})^2$ and can be derived as follows: $4(\boldsymbol{x}^T\boldsymbol{z})^2 = \sum_{i=1}^{i=n}\sum_{j=1}^{j=n} 4x_iz_ix_jz_j = \sum_{i=1}^{i=n}\sum_{j=1}^{j=n} 2x_ix_j * 2z_iz_j$. This leads us to the mapping $\phi_1(\boldsymbol{x}) = [2x_1x_1, 2x_1x_2, \ldots 2x_1x_n, 2x_2x_1, 2x_2x_2, \ldots 2x_2x_n, \ldots 2x_nx_1, 2x_nx_2 \ldots 2x_nx_n]^T$ which is a $n^2 \times 1$ mapping. $\phi_2(\boldsymbol{x})$ is the $n \times 1$ mapping $[\sqrt{12}x_1, \sqrt{12}x_2, \ldots \sqrt{12}x_n]$ representing the term $12\boldsymbol{x}^T\boldsymbol{z}$ and $\phi_3(\boldsymbol{x}) = \sqrt{12}$ representing the constant term in the kernel function. Now $\phi^T(\boldsymbol{x})\phi(\boldsymbol{x})$ can be seen to equal the given Kernel function and ϕ is of dimension $n^2 + n + 1$.

Q Answer the following questions with justifications.

(A) Consider the following design matrix, representing four sample points $X_i \in \mathbb{R}^2$:

$$\boldsymbol{X} = \begin{bmatrix} 4 & 1 \\ 2 & 3 \\ 5 & 4 \\ 1 & 0 \end{bmatrix}$$

(i) Compute the unit-length principal component directions of $\boldsymbol{X}$, and state which principal component you would choose if you were requested to choose just one. Show your computation.

[4 Marks]

(ii) Draw the principal component direction (as a line) and the projections of all four sample points onto the principal direction.

[3 Marks]

(iii) Reconstruct the data using rank-1 approximation.

[1 Marks]

(B) Assume that you are given the matrix A as below:

$$\boldsymbol{A} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

You are asked to use an iterative technique (e.g., power method) to obtain the eigenvector corresponding to the dominant eigenvalue. Your start with an initial value of $\mathbf{x} = [0, 0.5, 1]$ for the iterative technique. To which eigenvalue and eigenvector will the algorithm likely to converge? Given reasons for your answer.

[2 Marks]

Solution:

(A) (i)

x	y	(x - 3)	(y - 2)	$(x-3)^2$	$(y-2)^2$	(x - 3) (y - 2)
4	1	1	-1	1	1	-1
2	3	-1	1	1	1	-1
5	4	2	2	4	4	4
1	0	-2	-2	4	4	4
$\bar{\mathbf{x}} = \mathbf{3}$	$\bar{\mathbf{y}} = \mathbf{2}$	-	-	**10**	**10**	**6**

(1.5 marks)

Covariance Matrix:

$$\Sigma = \frac{1}{N} \begin{bmatrix} Cov(x,x) & Cov(x,y) \\ Cov(y,x) & Cov(y,y) \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 10 & 6 \\ 6 & 10 \end{bmatrix} = \begin{bmatrix} 2.5 & 1.5 \\ 1.5 & 2.5 \end{bmatrix}$$

(0.5 mark)

Eigenvalues:

$\det|\Sigma - \lambda I| = 0 => \lambda^2 - 5\lambda + 4 = 0 => \lambda_1 = 4, \lambda_2 = 1$

(1.0 mark)

Eigenvectors:

$$v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Unit Eigenvectors:

$$\hat{e}_1 = \begin{bmatrix} 0.71 \\ 0.71 \end{bmatrix}, \quad \hat{e}_2 = \begin{bmatrix} -0.71 \\ 0.71 \end{bmatrix}$$

(1.0 mark)

(ii) The one - dimensional subspace we are projecting onto is along the principal eigenvector:

$$\hat{e}_1 = \begin{bmatrix} 0.71 \\ 0.71 \end{bmatrix}$$

which corresponds to the direction of maximum variance in the data. The coordinates (in principal coordinate space) for each of the four sample points are projected as;

$$Y = \hat{e}_1^T X = \begin{bmatrix} 0.71 \\ 0.71 \end{bmatrix}^T X$$

and are given by:

$$(4,1) \rightarrow 3.55, (2,3) \rightarrow 3.55, (5,4) \rightarrow 6.39, (1,0) \rightarrow 0.71$$

(2.0 marks)

Projection:

`Projectionpoints.pdf`

(1 mark)

(iii) Projection for all data points can be written as follows:

$Z = X_{centered}.\hat{e}_1 =$

$$\begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 2 & 2 \\ 2 & -2 \end{bmatrix} \cdot \begin{bmatrix} 0.71 \\ 0.71 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 22 \\ -22 \end{bmatrix}$$

$X_{reconstructed} = Z.\hat{e}_1^T =$

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

which is the rank - 1 approximation
Note: In order to get the original data, Add mean to X reconstructed

(1 mark)

(B)

$X^{(1)} = AX_0 =$

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} . \begin{bmatrix} 0 \\ 0.5 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 4 \end{bmatrix} = 4 \begin{bmatrix} 0 \\ 0.5 \\ 1 \end{bmatrix} = \lambda_1 X_1$$

(1 marks)

$X^{(2)} = AX_1 =$

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix} . \begin{bmatrix} 0 \\ 0.5 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \\ 4 \end{bmatrix} = 4 \begin{bmatrix} 0 \\ 0.5 \\ 1 \end{bmatrix} = \lambda_2 X_2$$

(1 marks)

Here the method converges in second iteration.

The largest eigenvalue

$$\lambda = 4$$

and the corresponding eigenvector X =

$$\begin{bmatrix} 0 \\ 0.5 \\ 1 \end{bmatrix}$$

The converged eigenvalue is the same as the given initial vector. The given matrix is symmetric positive definite with the dominant eigenvalue repeated. Thus, any vector in the eigenspace of the dominant eigenvalue is a valid eigenvector.

Q Answer the following questions with justifications.

The block diagram below shows a system with input x, output L, and the computations.

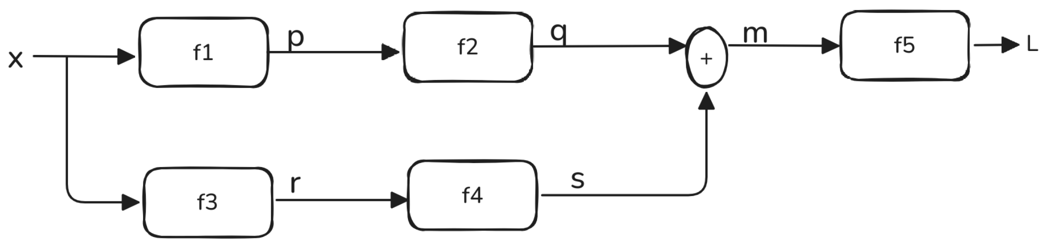

FIGURE 1

Here,

$$
\begin{aligned}
x &\in R^2 \\
f1 : p &= w_1^T x + b_1 \\
f2 : q &= \frac{1}{1+e^{-p}} \\
f3 : r &= w_2^T x + b_2 \\
f4 : s &= \frac{e^r - e^{-r}}{e^r + e^{-r}} \\
m &= q + s \\
L &= \frac{1}{2}(m - m_l)^2
\end{aligned}
$$

(A) Draw the computation graph for this system which can be used to compute the gradient of L w.r.t x using chain rule of partial derivatives.

[2 marks]

(B) show that:

$$\frac{\partial q}{\partial p} = q(1-q)$$

[2 marks]

(C) show that:

$$\frac{\partial s}{\partial r} = (1-s^2)$$

[2 marks]

(D) Write the expression for each of the below partial derivatives:

(i) $\frac{\partial L}{\partial r}$
(ii) $\frac{\partial L}{\partial w_1}$
(ii) $\frac{\partial L}{\partial w_2}$
(ii) $\frac{\partial L}{\partial x}$

[4 marks]

Solution

(A)

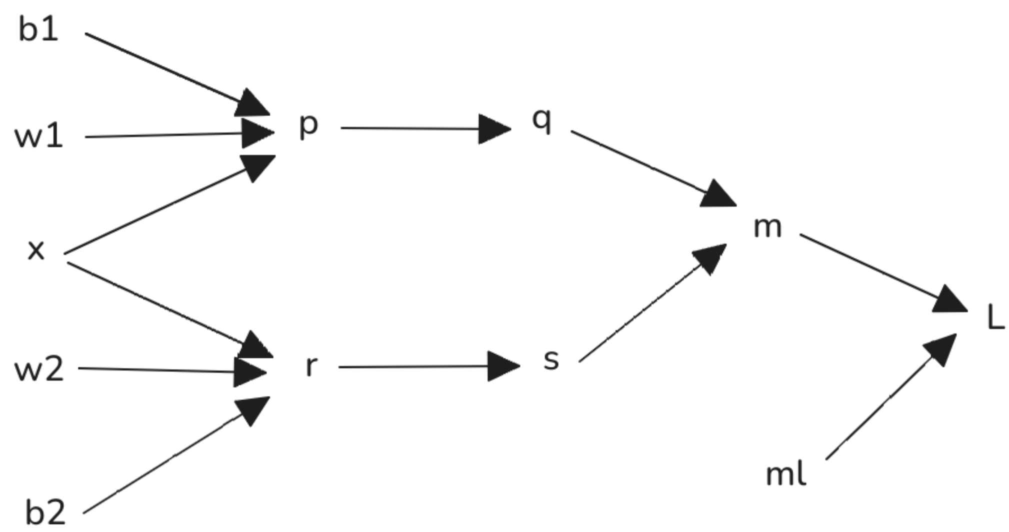

FIGURE 2

(B)

$$\begin{aligned} q &= \frac{1}{1+e^{-p}} \\ \frac{\partial q}{\partial p} &= \frac{e^{-p}}{(1+e^{-p})^2} \\ 1-q &= \frac{e^{-p}}{1+e^{-p}} \\ \frac{\partial q}{\partial p} &= q(1-q) \end{aligned}$$

(C)

$$\begin{aligned} s &= \frac{e^r - e^{-r}}{e^r + e^{-r}} \\ \frac{\partial s}{\partial r} &= \frac{(e^r + e^{-r})(e^r + e^{-r}) - (e^r - e^{-r})(e^r - e^{-r})}{(e^r + e^{-r})^2} \\ \frac{\partial s}{\partial r} &= 1 - \frac{(e^r - e^{-r})^2}{(e^r + e^{-r})^2} \\ \frac{\partial s}{\partial r} &= 1 - s^2 \end{aligned}$$

(D) (i)

$$\frac{\partial L}{\partial r} = \frac{\partial L}{\partial m}\frac{\partial m}{\partial s}\frac{\partial s}{\partial r} = (m - ml)(1 - s^2)$$

(ii)

$$\frac{\partial L}{\partial w1} = \frac{\partial L}{\partial m}\frac{\partial m}{\partial q}\frac{\partial q}{\partial p}\frac{\partial p}{\partial w1} = (m - ml)q(1 - q)x$$

(iii)

$$\frac{\partial L}{\partial w2} = \frac{\partial L}{\partial m}\frac{\partial m}{\partial s}\frac{\partial s}{\partial r}\frac{\partial r}{\partial w2} = (m - ml)(1 - s^2)x$$

(iv)

$$\frac{\partial L}{\partial x} = \frac{\partial L}{\partial p}\frac{\partial p}{\partial x} + \frac{\partial L}{\partial r}\frac{\partial r}{\partial x} =$$

$$(m - ml)q(1 - q)w1 + (m - ml)(1 - s^2)w2$$