Birla Institute of Technology and Science, Pilani

Work Integrated Learning Programmes Division

Cluster Programme - M.Tech. in Artificial Intelligence and Machine Learning.

Course Number	AIMLC ZC416
Course Name	Mathematical Foundations for Machine Learning
Nature of Exam	Closed Book
Weightage for grading	30%
Duration	120 minutes
Date of Exam	08/01/2025 (14:00 - 16:00)

# Pages	2
# Questions	8

Instructions

1. All questions are compulsory.
2. All parts of a question should be answered consecutively. Each answer should start from a fresh page.

---

(1) A data scientist is told that the characteristic equation of a data matrix he works with has the property that the roots are all distinct and real. The data scientist does not know the values of the roots of the characteristic equation but assumes that (a) the matrix is bound to be invertible and (b) also diagonalizable. Is the data scientist justified in making these assumptions about the matrix? Give appropriate reasons for your answer. [5 Marks]

Solution: Claim (a) is false. The reason for this is that one of the eigenvalues could be zero meaning that there is a non-zero vector in the nullspace of the matrix so that its rank is less than n. Claim (b) is true because the distinct eigenvalues ensure that each eigenspace is of dimension equal to 1. Thus there is a non-zero eigenvector associated with each eigenvalue. We also know that eigenvalues corresponding to distint eigenvalues are linearly independent, which means that we have n linearly independent eigenvectors and an invertible eigenvector matrix.

Marking Scheme: 2.5 Marks $\rightarrow$ Claim (a) , 2.5 Marks $\rightarrow$ Claim (b)

(2) Let A be a $m \times n$ matrix over $\mathbb{R}^2$. Define $N_A = \{x \in \mathbb{R}^n : Ax = 0\}$. N_A is called kernel of A and $\nu(A) = dim(N_A)$ as nullity of A.

(a) (i) Let

$$A = \begin{pmatrix} 1 & 2 & -1 \\ 2 & -1 & 3 \end{pmatrix}$$

Find N_A and $\nu(A)$ [2 Marks]

(b) (ii) Let

$$A = \begin{pmatrix} 2 & -1 & 3 \\ 4 & -2 & 6 \\ -6 & 3 & -9 \end{pmatrix}$$

Find N_A and $\nu(A)$. [3 Marks]

Solution: (i)
From

$$\begin{pmatrix} 1 & 2 & -1 \\ 2 & -1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

we find the system of linear equations

$$x_1 + 2x_2 - x_3 = 0 2x_1 - x_2 + 3x_3 = 0$$

Eliminating x_3 gives $x_1 = -x_2$. Also, $x_3 = -x_1$ Thus N_A is spanned by the vector

$$\begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$$

Hence $\nu(A)$=1.
(ii) From

$$\begin{pmatrix} 2 & -1 & 3 \\ 4 & -2 & 6 \\ -6 & 3 & -9 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

we find the system of linear equations

$$2x_1 - x_2 + 3x_3 = 0 4x_1 - 2x_2 + 6x_3 = 0 - 6x_1 + 3x_2 - 9x_3 = 0$$

All the three equations are same. Thus from the first equation i.e $2x_1 - x_2 + 3x_3 = 0$, we find that

$$\left\{ \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix} \right\}$$

he is a basis for N_A and $\nu(A) = 2$

(3) (i) Suppose if we arrive at a $m \times m$ matrix $\boldsymbol{A} = \boldsymbol{PDP}^T$ where $\boldsymbol{P}^{-1} = \boldsymbol{P}^T$ and $\boldsymbol{D} = \begin{bmatrix} l & 0 & \cdots & 0 \\ 0 & l^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & l^m \end{bmatrix}$ and $|l| \neq 0, 1$, then find a best rank 1 approximation of $\boldsymbol{A}^2$ with justification. [3 Marks]

(ii) A data analyst arrived at vectors $\boldsymbol{v}_1 = [1/\sqrt{(2)}, 1/\sqrt{(2)}]^T, \boldsymbol{v}_2 = [1/\sqrt{(2)}, -1/\sqrt{(2)}]^T$. Help the analyst to find a 2×2 matrix $\boldsymbol{A}$ such that $\boldsymbol{A}\boldsymbol{v}_i = -\boldsymbol{v}_i, \forall i = 1, 2$

Solution

(i) Now $\boldsymbol{A} = \boldsymbol{PDP}^T \Rightarrow \boldsymbol{AP} = \boldsymbol{PD}$.
$\Rightarrow \boldsymbol{AP}_i = l^i \boldsymbol{P_i},\ i = 1, \cdots m$ where $\boldsymbol{P}_i$ is the i^{th} column of $\boldsymbol{P}$.
Also $\boldsymbol{P}_i \neq \boldsymbol{0}$. Therefore, l^i is an eigenvalue of $\boldsymbol{A}$ and $\boldsymbol{P}_i$ is its corresponding eigenvector. [0.5 Marks]
$\Rightarrow l^{in}$ is an eigenvalue of $\boldsymbol{A}^n$ and $\boldsymbol{P}_i$ is its corresponding eigenvector for $i = 1, \cdots m$. [0.5 Marks]

Clearly $\boldsymbol{A}^T = (\boldsymbol{PDP}^T)^T = \boldsymbol{PDP}^T = \boldsymbol{A}$ as $\boldsymbol{D}$ is a diagonal matrix. Therefore, $\boldsymbol{A}^2$ is also a symmetric matrix and l^k are real.
So, $\boldsymbol{A}^2(\boldsymbol{A}^2)^T = (\boldsymbol{A}^2)^T\boldsymbol{A}^2 = \boldsymbol{A}^4$. $\Rightarrow l^{4i}$ is an eigenvalue and $\boldsymbol{P}_i$ is its corresponding eigenvector for $i = 1, \cdots m$ of matrices $\boldsymbol{A}^2(\boldsymbol{A}^2)^T$, $(\boldsymbol{A}^2)^T\boldsymbol{A}^2$.
$\Rightarrow l^{2i}$ is a singular value and $\boldsymbol{P}_i$ is its corresponding left and right singular vector for $i = 1, \cdots m$ of matrix $\boldsymbol{A}^2$. [1 Mark]
If $l > 1 \Rightarrow l^{2m}$ is the largest singular value and hence a best rank 1 approximation of $\boldsymbol{A}^2$ is $l^{2m}\boldsymbol{P}_m\boldsymbol{P}_m^T$. [0.5 Marks]
If $l < 1 \Rightarrow l$ is the largest singular value and hence a best rank 1 approximation of $\boldsymbol{A}^2$ is $l^2\boldsymbol{P}_1\boldsymbol{P}_1^T$. [0.5 Marks]
(Kindly award full marks for any other correct method.)

(ii) Now, clearly $\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$ forms an orthonormal basis for $\mathbb{R}^2$. Hence, $\{\boldsymbol{u}_1, \boldsymbol{u}_2\}$ forms an orthonormal basis for $\mathbb{R}^2$ where $\boldsymbol{u}_i = -\boldsymbol{v}_i, \forall i = 1, 2$. [1 Mark]
Define $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{I_2}\boldsymbol{V}^T$, where $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2], \boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2]$.
Then, $\boldsymbol{A}\boldsymbol{v}_i = \boldsymbol{u}_i = -\boldsymbol{v}_i, \forall i = 1, 2$. [1 Mark]

(4) Let $\mathbf{v}, \mathbf{w} \in \mathbb{R}^{n\times 1}$ be two non-zero vectors. Consider the identity matrix $\mathbf{I}_n \in \mathbb{R}^{n\times n}$.

(a) Consider the matrix $\mathbf{B} = \mathbf{I}_n - 2(\mathbf{v}\mathbf{v}^T)$. Prove or disprove whether this matrix $\mathbf{B}$ is an orthogonal matrix for any $\mathbf{v}$. Also prove/disprove whether $\mathbf{B}$ is a symmetric matrix for any $\mathbf{v}$. [1.5 Marks]

(b) Assume that the Euclidean norm of $\mathbf{w}$ is equal to 1. Consider the matrix $\mathbf{C} = \mathbf{I}_n - 2(\mathbf{w}\mathbf{w}^T)$. Prove or disprove whether this matrix $\mathbf{C}$ is an orthogonal matrix. [1.5 Marks]

(c) Show that when the Euclidean norm of $\mathbf{v}$ is equal to 1, the rank of $\mathbf{I}_n - \mathbf{v}\mathbf{v}^T$ is less than n. [2 Marks]

Solution:

(a) Calculate $\mathbf{B}^T\mathbf{B} = (\mathbf{I}_n - 2(\mathbf{v}\mathbf{v}^T))^T(\mathbf{I}_n - 2(\mathbf{v}\mathbf{v}^T))$=$I - (v^Tv - 1)(4vv^T)$. Hence B is not orthogonal matrix. Now to check for symmetric property: $B^T = (\mathbf{I}_n - 2(\mathbf{v}\mathbf{v}^T))^T = (\mathbf{I}_n^T - 2(\mathbf{v}\mathbf{v}^T)^T)$=$(\mathbf{I}_n - 2(\mathbf{v}\mathbf{v}^T)) = \mathbf{B}$.Hence B is symmetric matrix. (1.5 marks)

(b) Calculate $\mathbf{C}^T\mathbf{C} = (\mathbf{I}_n - 2(\mathbf{w}\mathbf{w}^T))^T(\mathbf{I}_n - 2(\mathbf{w}\mathbf{w}^T))$=$I - (w^Tw - 1)(4ww^T)$. But recall that $\|w\|_2^2 = 1$. We know that $w^Tw = \|w\|_2^2 = 1$. Substituting this in previous expression we get: $\mathbf{C}^T\mathbf{C} = I - (1-1)(4ww^T) = I$. (0.75 marks)
Similarily we can show that $\mathbf{C}\mathbf{C}^T = I$ (0.75 marks). Hence C is orthogonal.

(c) We can see that $(\mathbf{I}_n - \mathbf{v}\mathbf{v}^T)\mathbf{v} = \mathbf{v} - (\mathbf{v}^T\mathbf{v})\mathbf{v} = \mathbf{v} - \mathbf{v} = \mathbf{0}$. Thus we have a non-zero vector $\boldsymbol{v}$ in the nullspace of $\mathbf{I}_n - \mathbf{v}\mathbf{v}^T$. (1 Mark)
By the rank-nullity theorem this means that the rank of $\mathbf{I}_n - \mathbf{v}\mathbf{v}^T$ is less than n. (1 Mark)

(5) The profit of a software company on a project is given by

$$x_1^3 + x_2^3 + 2x_1^2 + 4x_2^2 + 6$$

where x_1 and x_2 denote, respectively, the salary of employees and the general expenditure. Find the Hessian matrix and check whether it is positive definite or negative definite at each stationary point. Hence, find the values of x_1 and x_2 to maximize the profit.

Solution

$$\frac{\partial f}{\partial x_1} = 0 \implies 3x_1^2 + 4x_1 = 0 \implies x_1 = 0, -4/3$$
$$\frac{\partial f}{\partial x_2} = 0 \implies 3x_2^2 + 8x_2 = 0 \implies x_2 = 0, -8/3$$
$$\frac{\partial^2 f}{\partial x_1^2} = 6x_1 + 4$$
$$\frac{\partial^2 f}{\partial x_2^2} = 6x_2 + 8$$
$$\frac{\partial^2 f}{\partial x_1 \partial x_2} = 0$$

Hence, the Hessian matrix is

$$\mathbb{J} = \begin{bmatrix} 6x_1 + 4 & 0 \\ 0 & 6x_2 + 8 \end{bmatrix}$$

$J_1 = |6x_1 + 4|$ and $J_2 = \begin{vmatrix} 6x_1 + 4 & 0 \\ 0 & 6x_2 + 8 \end{vmatrix}$

At (0, 0) $J_1 = 4 > 0, J_2 = 32 > 0$, therefore, $\mathbb{J}$ is positive definite. Therefore, $f(x)$ has a relative minimum at (0, 0)

At (0, -8/3) $\mathbb{J}$ is indefinite, therefore (0, -8/3) is a saddle point.

At (-4/3, 0), $\mathbb{J}$ is again indefinite, so (-4/3, 0) is a saddle point.

At (-4/3, -8/3) $\mathbb{J}$ is negative definite, therefore, f has a relative maximum at (-4/3, -8/3).

(6) Let $A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \in \mathbb{R}^{2\times 3}$. Consider the function $f : \mathbb{R}^{2\times 3} \to \mathbb{R}^{3\times 3}$ given by $f(A) = A^T A = K$. Calculate the gradient of K with respect to A in matrix form.

Solution

$$K = A^T A = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

(1M)
$$= \begin{bmatrix} {a_{11}}^2 + {a_{21}}^2 & a_{11}a_{12} + a_{21}a_{22} & a_{11}a_{13} + a_{21}a_{23} \\ a_{12}a_{11} + a_{22}a_{21} & {a_{12}}^2 + {a_{22}}^2 & a_{12}a_{13} + a_{22}a_{23} \\ a_{13}a_{11} + a_{23}a_{21} & a_{13}a_{12} + a_{23}a_{22} & {a_{13}}^2 + {a_{23}}^2 \end{bmatrix}$$

The gradient of K with respect to A is $9x6$ matrix given by

$$\frac{dK}{dA} = \begin{bmatrix} \frac{\partial K_{11}}{\partial A_{11}} & \frac{\partial K_{11}}{\partial A_{12}} & \frac{\partial K_{11}}{\partial A_{13}} & \frac{\partial K_{11}}{\partial A_{21}} & \frac{\partial K_{11}}{\partial A_{22}} & \frac{\partial K_{11}}{\partial A_{23}} \\ \frac{\partial K_{12}}{\partial A_{11}} & \frac{\partial K_{12}}{\partial A_{12}} & \frac{\partial K_{12}}{\partial A_{13}} & \frac{\partial K_{12}}{\partial A_{21}} & \frac{\partial K_{12}}{\partial A_{22}} & \frac{\partial K_{12}}{\partial A_{23}} \\ \frac{\partial K_{13}}{\partial A_{11}} & \frac{\partial K_{13}}{\partial A_{12}} & \frac{\partial K_{13}}{\partial A_{13}} & \frac{\partial K_{13}}{\partial A_{21}} & \frac{\partial K_{13}}{\partial A_{22}} & \frac{\partial K_{13}}{\partial A_{23}} \\ \frac{\partial K_{21}}{\partial A_{11}} & \frac{\partial K_{21}}{\partial A_{12}} & \frac{\partial K_{21}}{\partial A_{13}} & \frac{\partial K_{21}}{\partial A_{21}} & \frac{\partial K_{21}}{\partial A_{22}} & \frac{\partial K_{21}}{\partial A_{23}} \\ \frac{\partial K_{22}}{\partial A_{11}} & \frac{\partial K_{22}}{\partial A_{12}} & \frac{\partial K_{22}}{\partial A_{13}} & \frac{\partial K_{22}}{\partial A_{21}} & \frac{\partial K_{22}}{\partial A_{22}} & \frac{\partial K_{22}}{\partial A_{23}} \\ \frac{\partial K_{23}}{\partial A_{11}} & \frac{\partial K_{23}}{\partial A_{12}} & \frac{\partial K_{23}}{\partial A_{13}} & \frac{\partial K_{23}}{\partial A_{21}} & \frac{\partial K_{23}}{\partial A_{22}} & \frac{\partial K_{23}}{\partial A_{23}} \\ \frac{\partial K_{31}}{\partial A_{11}} & \frac{\partial K_{31}}{\partial A_{12}} & \frac{\partial K_{31}}{\partial A_{13}} & \frac{\partial K_{31}}{\partial A_{21}} & \frac{\partial K_{31}}{\partial A_{22}} & \frac{\partial K_{31}}{\partial A_{23}} \\ \frac{\partial K_{32}}{\partial A_{11}} & \frac{\partial K_{32}}{\partial A_{12}} & \frac{\partial K_{32}}{\partial A_{13}} & \frac{\partial K_{32}}{\partial A_{21}} & \frac{\partial K_{32}}{\partial A_{22}} & \frac{\partial K_{32}}{\partial A_{23}} \\ \frac{\partial K_{33}}{\partial A_{11}} & \frac{\partial K_{33}}{\partial A_{12}} & \frac{\partial K_{33}}{\partial A_{13}} & \frac{\partial K_{33}}{\partial A_{21}} & \frac{\partial K_{33}}{\partial A_{22}} & \frac{\partial K_{33}}{\partial A_{23}} \end{bmatrix} \text{[2M}$$

(or) one can mention the formula as described in the text book and the same 2 marks will be awarded.

$\frac{dK}{dA} = \left[\frac{\partial K_{pq}}{\partial A_{ij}}\right] = [\partial_{pqij}]$

Where $\partial_{pqij} = \begin{cases} A_{\text{iq}} \text{ if } j = p, & p \neq q \\ A_{\text{ip}} \text{ if } j = q, & p \neq q \\ 2A_{\text{iq}} \text{ if } j = p, & p = q \\ 0 \text{ otherwise} \end{cases}$

Hence

$$\frac{dK}{dA} = \begin{bmatrix} 2a_{11} & 0 & 0 & 2a_{21} & 0 & 0 \\ a_{12} & a_{11} & 0 & a_{22} & a_{21} & 0 \\ a_{13} & 0 & a_{11} & a_{23} & 0 & a_{21} \\ a_{12} & a_{11} & 0 & a_{22} & a_{21} & 0 \\ 0 & 2a_{12} & 0 & 0 & 2a_{22} & 0 \\ 0 & a_{13} & a_{12} & 0 & a_{23} & a_{22} \\ a_{13} & 0 & a_{11} & a_{23} & 0 & a_{21} \\ 0 & a_{13} & a_{12} & 0 & a_{23} & a_{22} \\ 0 & 0 & 2a_{13} & 0 & 0 & 2a_{23} \end{bmatrix}$$

[2M] [The final answer must be in the matrix form]