Indian Institute of Technology (IIT) Roorkee
Department of Electrical Engineering

Set-0101

EEC-351: Fundamentals of AI/ML

Mid-Term Examination (Autumn 2026–27)

Instructors: Parikshit Pareek

Timing: 90 Mins **Date: Sep 10, 2026** **Time: 02:15 PM – 03:45 PM** **Max. Marks: 45**

Instructions

- Answer all questions. Each question must begin on a **NEW PAGE**. Write the Set name on the top right corner of the main answer sheet.
- Provide complete justifications. Mention assumptions and conditions wherever needed. Unjustified answers may receive zero marks.
- Throughout, $\|\cdot\|$ without a subscript means the ℓ_2-norm, $\kappa_2(\cdot)$ the ℓ_2 condition number, and $\mathcal{R}(\cdot)$, $\mathcal{N}(\cdot)$ the range (column space) and null space.
- All the question carry **[-1]** for unnecessary long answers, particularly story telling, even if not mentioned in the question itself.

[Pre-Condition Your Mind & Remember in ETE that This MTE is `Easier` Than Last Year.]

1. Let $A = \begin{bmatrix} 100 & 0 \\ 0 & 0.01 \end{bmatrix}$ and $B = \frac{1}{500}\begin{bmatrix} 3 & 4 \\ -4 & 3 \end{bmatrix}$.

   (a) One of them is nearly singular: which one, and what does "nearly singular" mean geometrically? Support your verdict with a computation. **[Three lines max, [-1] else.]** **[1+1]**

   (b) Let $A \in \mathbb{R}^{n\times n}$ is invertible and scalar $c \neq 0$, compare $\det(cA)$ with $\det(A)$, and $\kappa_2(cA)$ with $\kappa_2(A)$. What does this say about det as a detector of near-singularity? **[1]**

   (c) For A above, let $b = (100,\ 0.01)^\top$. Solve $Ax = b$. Now perturb to $b' = (100,\ 0.02)^\top$ and solve again. Compute $\|\Delta b\| \,/\, \|b\|$ and $\|\Delta x\| \,/\, \|x\|$. **Derive** the general upper bound on $\|\Delta x\| \,/\, \|x\|$ in terms of condition number. **[1+2]**

2. Let $X = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 4 \\ 1 & 3 & 6 \end{bmatrix}$ and $y = (1,\ 2,\ 2)^\top$.

   (a) State $\text{rank}(X)$, and if it is full rank. Is $X^\top X$ invertible? **[Number & Yes/No only]** **[1]**

   (b) Compute the orthogonal projection of y onto $\mathcal{R}(X)$ & verify that residual $r \perp \mathcal{R}(X)$. **[3]**

   (c) Mathematically describe the set of *all* $\hat{w} \in \mathbb{R}^3$ satisfying $X\hat{w} = \hat{y}$. **[2]**

   (d) Prove without solving that $\forall \lambda > 0$ the ridge system $(X^\top X + \lambda I)\hat{w} = X^\top y$ has a solution for $\hat{w}$. **[1]**

3. Let $X \in \mathbb{R}^{n\times d}$ with $n \geq d$ and $\text{rank}(X) = d$, with singular values $\sigma_1 \geq \cdots \geq \sigma_d > 0$, so that $\kappa_2(X) = \sigma_1/\sigma_d$. Prove that $\kappa_2(X^\top X) = \{\kappa_2(X)\}^2$. **[2]**

4. Two IITR UGs (*aka ML Researchers*) independently estimate the covariance matrix of three features (X_1, X_2, X_3) from the *same* dataset and report

$$\Sigma_A = \begin{bmatrix} 4 & 3 & 0 \\ 3 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \Sigma_B = \begin{bmatrix} 4 & 2 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

   Which one is correct and why? Give proof of the property you use to verify the correctness. **[1+2]**

5. A vector-search engine ranks stored embeddings $x^{(j)} \in \mathbb{R}^d$ against a query 'q' by a "distance" $\text{dist}(x, q) = f(x - q)$. For the ranking to behave like a distance, f must be a norm. Four candidates are proposed:

   (i) $f_1(x) = \|x\|_2^2$;

   (ii) $f_2(x) = \sqrt{x^\top A x}$ with $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$, $d = 2$ ("a learned metric");

(iii) $f_3(x) = \sqrt{x^\top A x}$ with $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$, $d = 2$; ("a differently learned metric");

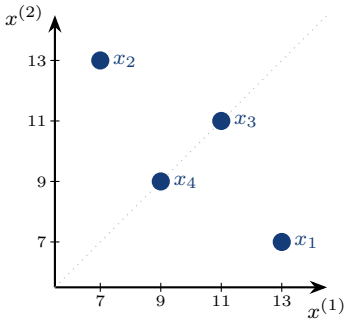
(iv) $f_4(x) = \max_i |x_i| + \min_i |x_i|$, $d = 3$.

(a) For each candidate, **show** whether it is a norm on the stated space. If it is not, name one axiom that fails and give a numerical counterexample. **[Mathematical Argument, only]** **[1+1+1+1]**

(b) Does your verdict on f_4 change if $d = 2$? Justify in one line. **[Unwanted theory [-1]]** **[1]**

(c) Prove: if $A \in \mathbb{R}^{d\times d}$ is symmetric positive definite, then $f_A(x) = \sqrt{x^\top A x}$ is a norm on $\mathbb{R}^d$. **[1]**

6. Four two-dimensional data points are plotted alongside. One of the IITR UG, runs PCA but forgets to center the data (*Ya, he was not using default PCA packages! Brave Guy!*).



(a) Compute the matrix $M = \frac{1}{n}\sum_i x_i x_i^\top$ that the intern eigendecomposes, its top eigenvector v_M, and the "variance explained" ratio the intern reports. **[2]**

(b) Now center the data. Compute the covariance matrix C, the true first principal component v_C, and the true explained-variance ratio. **[2]**

7. The *Skynet* (*Terminator Fame*) compresses a grayscale image $X \in \mathbb{R}^{m\times n}$ by keeping only the top singular triple i.e. $X_1 = \sigma_1 u_1 v_1^\top$, and claims the file now stores "one column, one row, and one number." Show that for every $x \in \mathbb{R}^n$ the vector $X_1 x$ is a scalar multiple of u_1. Also, determine $\mathcal{R}(X_1)$ and prove $\operatorname{rank}(X_1) = 1$ *exactly*. **[2]**

8. Let all norms ($\|\cdot\|$) are convex operators. For each function below, decide whether it is convex or non-convex, and prove your verdict.

(a) $f(w) = 3\|w\|_1 + 2\|Xw - y\|_2 + |a|\,\|w - y\|_2 \quad \forall a \in [-1, 1]$, X `is Indefinite Matrix` **[1]**

(b) Let $A \in \mathbb{R}^{m\times n}$ be arbitrary and $g(w) = w^\top A^\top A\, w - (214) b^\top A w$. **[1]**

(c) $h(w) = \max\{\, \|Aw - b\|_2^2,\, \|Cw - d\|_2^2 \,\} + \lambda \|w\|_2^2$. **[1]**

9. Let, $A \in \mathbb{R}^{3\times 3}$ *stretch-free* if $\|Ax\|_2 = \|x\|_2$ for every $x \in \mathbb{R}^3$.

(a) Design a stretch-free, non-identity matrix A, and prove that your matrix is stretch-free. **[2]**

(b) Consider a stretch-free matrix R. Find $\kappa_2(R) =?$, Show that any *real* eigenvalue of R is ± 1. **[1]**

10. During commissioning of a protection relay (*Yes! Borrowed from Morning Exam*), a test window records five readings (in mA) : $\mathcal{S} = \{1, 2, 3, 99\}$. The 99 is a suspected switching transient. The report must publish one representative value '**c**', and three engineers propose three criteria. The *statistician* wants the total squared error. The *field engineer* wants the total absolute error. The *safety guy* wants the worst case. Find the minimizer of each criterion for $\mathcal{S}$. **[2]**

11. Gradient descent $w_{k+1} = w_k - \eta \nabla f(w_k)$ is run on

$$f(w) = \tfrac{1}{2} w^\top A w - b^\top w, \qquad A = \begin{bmatrix} 6 & 2 \\ 2 & 3 \end{bmatrix}, \quad b = \begin{bmatrix} 8 \\ 5 \end{bmatrix}.$$

(a) Find the minimizer $w^\star$ and the eigenpairs of A. **[1]**

(b) Let $e_k = w_k - w^\star$. Derive the exact condition on η under which $e_k \to 0$ from *every* starting point w_0, and evaluate it numerically for this A. **[3]**

12. A zero-mean signal is modelled as $X \sim \text{Uniform}(-1, 1)$, and a monitoring routine also has $Y = X^2$.

(a) Compute $\mathbb{E}[X]$, $\mathbb{E}[X^2]$, $\mathbb{E}[X^3]$ and show $\text{Cov}[X, Y] = 0$. **[1]**

(b) Are X, Y dependent or independent? Why? **[1]**

(d) Now let (X_1, X_2) be jointly Gaussian with mean 0 and $\Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$.

Are (X_1, X_2) independent? Why? **[1]**

[This is The End! Hold Your Breath & Count to Ten– Sky`Might`Fall !]