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

Set-A

EEC-351: Fundamentals of AI/ML

End-Term Examination (Autumn 2025–26)

Timing: 3 hrs | Date: Nov 17, 2025 Time: 9:30 AM - 12:30 PM | Max. Marks: 40

Instructions

- Answer all questions. Each question must begin on a **new page**.
- Provide complete justifications for your answers. Partial/unjustified answers may not receive full credit.
- Mention assumptions and conditions clearly wherever necessary.

---

1. Consider a dataset $\{(x_n, y_n)\}_{n=1}^N$ and a nonlinear transformation $z_n = \Phi(x_n)$. The PLA is run on the transformed data $\{z_n, y_n\}$ and is observed to converge in T updates. Assume:

   - $\|z_n\| \leq R$ for all n; • the final PLA weight vector satisfies $\|w\| = W$;
   - the margin of this separator is $\gamma = \min_n y_n\, w^\top z_n$ (value unknown).

   Now, before running PLA again, you apply a second linear mapping $u_n = \Psi(z_n)$, where Ψ has singular values in the interval $[s_{\min},\, s_{\max}]$. Let T' be the number of PLA updates required after the second mapping.

   (a) Derive upper and lower bounds on T' in terms of $s_{\min}, s_{\max}$ and T. **[4]**

   (b) Give a simple condition (in terms of $s_{\min}$ and $s_{\max}$) under which PLA is guaranteed to make fewer updates after the mapping, i.e. $T' < T$. **[4]**

2. Let x is some integer in the set $X = \{1, 2, \ldots, 50, 51, 52\}$, and where each hypothesis $h \in \mathcal{H}$ is an interval of the form $b \leq x \leq a$, with b and a as any integers between 1 and 52 (inclusive), so long as $b \leq a$. A hypothesis $b \leq x \leq a$ labels instance x positive if x falls into the interval defined by a and b, and labels the instance negative otherwise.

   (a) How many distinct hypotheses are there in such $\mathcal{H}$?. (No explanation required) **[0.5]**

   (b) Suppose we draw N independent examples uniformly from X with $\mathcal{H}$ hypothesis space. Using Hoeffding's inequality: If $\epsilon = 0.05$, calculate the minimum number of samples N needed to ensure that the confidence is at least 95%. **[0.5]**

   (c) If instead of Hoeffding's, we use Chebyshev inequality then what will be N? What is making N with Chebyshev inequality larger or smaller compared to Hoeffding's? **[1]**

3. Derive the closed-form (analytical) expression of weights update in Linear Regression model, define all variables and their sizes and state all assumptions. **[4]**

4. Write pseudo code of PLA and Pocket Algorithms. Can PLA converge? State the condition and prove if it does. **[4+4]**

5. Answer the following questions. **[3]**

   (a) Given $\mathbf{w} = (2, -1, 0.5)$ and input $\mathbf{x} = (1, 2, 1)$, the perceptron prediction $\text{sign}(\mathbf{w}^T\mathbf{x})$ is:

   (b) Current weight: $\mathbf{w} = (0,\ 1,\ -2)$. Misclassified point: $\mathbf{x} = (1,\ -2,\ 3)$, $y = +1$. What is the updated weight after one PLA update?

   (c) Given $h(\mathbf{x}) = w_0 + w_1 x$ with $w_0 = 1.2$ and $w_1 = 0.8$, the predicted output for $x = 4$ is:

6. Hoeffding's inequality provides concentration bounds for sums of independent bounded random variables. Explain mathematically why the independence assumption is essential for Hoeffding's inequality, and what breaks down when the X_i's are not independent. Also state Azuma's Inequality and condition under which it is valid. **[1+4]**

7. Consider a binary classifier that applies the following quadratic feature transformation to an input vector $\mathbf{x} = (x_1, x_2)$:

$$\mathbf{z} = \begin{bmatrix} z_0 \\ z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 1 \\ x_1^2 + x_2^2 \\ x_1 x_2 \end{bmatrix}.$$

   The PLA learns a linear classifier in the transformed space: $f(\mathbf{z}) = w_0 + w_1 z_1 + w_2 z_2 = 0$.

(a) Derive the decision boundary equation in the original x-space in terms of x_1 and x_2. Simplify the expression as far as possible. **[1]**

(b) For the following weight vectors (w_0, w_1, w_2), identify the geometric shape of the resulting decision boundary in x-space (circle, ellipse, parabola, hyperbola, pair of lines, etc.): **[2]**

1. $w_0 = -9,\ w_1 = 1,\ w_2 = 0$
2. $w_0 = -16,\ w_1 = 1,\ w_2 = -1$
3. $w_0 = 0,\ w_1 = 0,\ w_2 = 1$
4. $w_0 = 0,\ w_1 = 1,\ w_2 = 1$

8. You are given a linear regression model: $\hat{y} = wx + b$ with the following training dataset:

i	x_i	y_i
1	1	4
2	2	7
3	3	10

The loss function is: $L(w, b) = \frac{1}{2m} \sum_{i=1}^{m} (wx_i + b - y_i)^2$. Assume the initial parameters: $w = 0, b = 0$, and learning rate: $\eta = 0.1$. Answer the following:

(a) Compute the initial cost $L(w, b)$ at $w = 0, b = 0$. Perform one full step of Batch Gradient Descent (using all 3 samples). Compute updated values of w and b. **[1]**

(b) Perform two sequential steps of Stochastic Gradient Descent using samples in order $(1 \to 2)$. Start again from $w = 0, b = 0$. Show updated values after each step. **[2]**

9. For linear regression, the out-of-sample error is $E_{\text{out}}(h) = \mathbb{E}\big[(h(x) - y)^2\big]$. Show that among all hypotheses, the one that minimizes E_{out} is given by

$$h^*(x) = \mathbb{E}[\, y \mid x \,].$$

The function h^* can be treated as a deterministic target function, in which case we can write $y = h^*(x) + \epsilon(x)$, where $\epsilon(x)$ is an (input–dependent) noise variable. Show that $\epsilon(x)$ has expected value zero. **[4]**

End of Paper