Birla Institute of Technology & Science, Pilani
Work Integrated Learning Programmes Division
Second Semester 2024-2025

Comprehensive Examination
EC-3 Regular

Course No.	: AIMLC ZC416
Course Title	: Mathematical Foundations for Machine Learning
Nature of Exam	: Closed Book / Open Book (As per Course Handout)
Weightage	: 40%
Duration	: 2 Hours 30 mts
Date of Exam	: 06-09-2025, AN

No. of Pages = 2
No. of Questions = 6

Note to Students:

1. Please follow all the *Instructions to Candidates* given on the cover page of the answer book.
2. All parts of a question should be answered consecutively. Each answer should start from a fresh page.
3. Assumptions made if any, should be stated clearly at the beginning of your answer.

Q1. Consider the matrix $A = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 2 & -1 & 0 & 1 \\ 0 & -2 & -2 & -4 \\ 0 & 1 & 1 & 2 \end{bmatrix}$.

(a) Find the rank of the matrix A. [3M]

(b) Let W be the vector space spanned by the columns of A. Find a basis and dimension of W. [2M]

(c) Show that the set $U = \left\{ r \begin{bmatrix} 1 \\ 1 \\ -2 \\ 1 \end{bmatrix} : r \text{ is a real number} \right\}$ is a subspace of W. [2M]

Q2. Let $A = \begin{bmatrix} -2 & 0 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}$.

(a) Find the singular value decomposition of $A = U\Sigma V^T$. [4M]

(b) Suppose if you consider the matrix A as data matrix (data is sample data), what are eigen values of the covariance matrix of A. Also determine the direction along which the variance is maximum. [3M]

Q3. (a) Let $w(x, y, z) = xy + yz + xz$ and $x(u, v) = u + v,\ y(u, v) = u^2 - v,$ and $z(u, v) = u - v^2$. Using the chain rule, find $\frac{\partial w}{\partial u}$ and $\frac{\partial w}{\partial v}$. [3M]

(b) Assume that a signal tower is placed at (0, 0) and the cost to place a receiver at a point (x, y) is $3(x^2 + y^2)$. The region where we want to place the receiver R = $\{(x,y): x+y-1 \geq 0 \text{ and } x+y-3 \leq 0\}$.

(i) Formulate the constrained optimization problem to find the point at which the cost to place the receiver in the region R is minimum. [1M]

(ii) Formulate a Lagrangean dual problem for the primal problem you formulated in (i). [2M]

Q4. Consider the function of two variables $f(x, y) = (x - 1)^2 + (y - 1)^2$. Suppose you apply gradient descent algorithm with momentum with initial values $x_0 = 1.5, y_0 = 1.5$, the learning parameter α = 0.05 and the momentum parameter β = 0.8 to find the point at which $f(x, y)$ attains its global minimum. Find (x_1, y_1), (x_2, y_2) using first two iterations of gradient descent algorithm with momentum. Also find (a_1, b_1), (a_2, b_2) using first two iterations of gradient descent algorithm using the same initial values $x_0 = 1.5, y_0 = 1.5$ and the learning parameter α = 0.05. Conclude whether there is any advantage of using the gradient descent algorithm with momentum instead of just using the gradient descent algorithm? [7M]

Q5. Consider the following soft margin SVM problem:
You are given a 2-dimensional data set and their corresponding labels:
$x_1 = (0,0),\ y_1 = -1$
$x_2 = (2,2),\ y_2 = 1$
$x_3 = (2,0),\ y_3 = -1$
$x_4 = (0,2),\ y_4 = 1$

Suppose we use the soft margin SVM objective with C = 1:
$\min \frac{1}{2}||w||^2 + C\sum_{i=1}^{4} \max(0,\ 1 - y_i(w^T x_i + b))$.

Suppose the following two classifiers are proposed:
Model A: $w = (1,1)$ *and* $b = -1.5$
Model B: $w = (1,-1)$*and* $b = 0$

Compute the following:
(a) For each model compute the margin term $\frac{1}{2}||w||^2$. [1M]
(b) For each data point, compute the hinge loss $l_i = \max(0,\ 1 - y_i(w^T x_i + b))$, then sum them. [3M]
(c) Compute the total objective values for Model A and Model B, and state which model is better. [1M]
(d) For the better model you determined in (c), which points have margin errors and which points are correctly classified. [2M]

Q6. Suppose you are given a 1-dimensional data that is not linearly separable:
Inputs x = [-1.5, -0.5, 0.5, 1.5] and their corresponding Labels y = [1, -1, -1, 1].
Suppose you consider the feature map $\varphi(x) = [1,\ \sqrt{2}\,x, x^2]$.
(a) Show that the kernel function corresponding to the feature map φ is given by $k(x,z) = (1 + xz)^2$. [3M]
(b) Compute the 4x4 kernel matrix $K_{ij} = K(x_i, x_j)$. [3M]
