

Birla Institute of Technology and Science, Pilani
Work Integrated Learning Programmes Division
Cluster Programme - M.Tech. in AIML

II Semester 2022-23

Course Number	AIMLC ZC416
Course Name	Mathematical Foundation for Machine Learning
Nature of Exam	Open Book
Weightage for grading	40%
Duration	
Date of Exam	

# Pages	4
# Questions	4

Instructions

1. All questions are compulsory
 2. Questions are to be answered in the order in which they appear in this paper and in the page numbers mentioned before each of them.
-

Q1 Answer the following questions with justifications.

a A data science intern arrived at a loss function given by

$$f(x) = 3x^4 - 20x^3 + 36x^2 + 10.$$

- i) Help the intern to find the stationary points and classify and hence the global minima. [3 marks]
- ii) Suggest the intern whether $x = 0.5$ or $x = 3.5$ is a better initial condition to find global minima using simple gradient descent method with reasons. [1 mark]

b A Professor gave a 2 dimensional data matrix $\mathbf{X} = \begin{pmatrix} 2 & -1 & -2 & 1 \\ -4 & 2 & 4 & -2 \end{pmatrix}$

for dimension reduction. Student 1 suggested that the direction of maximum variance is $[1, 0]^T$. Student 2 suggested that the direction of maximum variance is $[0, 1]^T$. Student 3 claimed that both student 1 and 2 are wrong. Find the correct v that gives the direction of maximum variance and hence decide on which student is correct. [2 marks]

c The set of interest for a data scientist was

$$\mathbf{M} = \{(x, y) \in \mathbb{R}^2 | x^2 + y^2 \leq 5, x + 2y = 4, x \geq 0, y \geq 0\}.$$

- i) Formulate a constrained optimization problem to find a point in \mathbf{M} nearest to the origin $(0, 0)$. [1 mark]
- ii) Write the Lagrangian function for the above problem. [1 mark]
- iii) Verify and find the values of Lagrangian multipliers such that the point $(\frac{4}{5}, \frac{8}{5})$ satisfies the KKT conditions for the above problem. Justify your steps. [2 marks]

Q2 Answer the following questions with justifications.

- i) Find the rank of a $A_{n \times n}$ matrix whose jk^{th} entry is $a_{jk} = j + k - 1$. Would it change if $a_{jk} = j + k - \alpha$, where α is a positive integer? Justify. [2 marks]
- ii) Prove or disprove the statement that the dimension of a vector space is independent of the field over which it is defined. [2 marks]
- iii) Prove that if v_1, v_2, v_3 are elements of a vector space V over a field F such that $v_1 + v_2 + v_3 = 0$, then $\{v_1, v_2\}$ spans the same subspace as $\{v_2, v_3\}$ [2 marks]
- iv) Two students A and B perform RREF on a given matrix X and obtain the same matrix \tilde{X} . Student C says both A and B cannot have the same RREF whereas D says that it is possible. A senior student E says that both A and B could be correct and it is actually possible to get X from \tilde{X} , provided none of the elements in X is zero. Who all are correct and why? [2 marks]
- v) Given distinct eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$, is it possible to generate an orthogonal matrix of order n ? Justify your answer. [2 marks]

Q3 Answer the following questions with justifications.

- (1) i) Let $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^n$. Consider the following function:

$$d(x, y) = \sum_{i=1}^n |x_i - y_i|$$

Prove or disprove that $d(x, y)$ is a valid distance function by verifying all the properties. [2 marks]

- ii) Consider the set $\mathcal{C} = \{x | Ax = b\}$ where $A \in \mathbb{R}^{n \times n}$ is square matrix and $b \in \mathbb{R}^n$ and \mathcal{C} here represents all possible solutions of a linear system $Ax = b$. Prove or disprove that \mathcal{C} is a convex set. [1 mark]
- (2) A data scientist came across a dataset x_1, x_2, \dots, x_N where $x_i \in \mathbb{R}^{1024}$. Here $N = 20$. He wants to find the eigenvalues of a new matrix defined as follows

$$S = \sum_{i=1}^N x_i x_i^T$$

Observe that here $S \in \mathbb{R}^{1024 \times 1024}$. The data scientist has access to a piece of octave code containing a function $[P, D] = eig(A)$ which returns eigenvalues and eigenvectors of a square symmetric matrix A as long as it has less than 32 rows. (Note that $eig(A)$ returns matrices $P \in \mathbb{R}^{n \times n}$ of eigenvectors and a diagonal matrix $D \in \mathbb{R}^{n \times n}$ of eigenvalues.)

- (a) How will you use the $eig(A)$ function to find eigenvalues of S by overcoming the fact that S has 1000 rows. Give a mathematical reasoning of how this can be achieved. [1.5 marks]
- (b) How will you use the $eig(A)$ function to find eigenvectors of S by overcoming the fact that S has 1000 rows. Give a mathematical reasoning of how this can be achieved. [1.5 marks]
- (3) Consider the function $g(z) = \frac{1}{2}\|Az - b\|_2^2 + \frac{1}{2}\|z\|_2^2 + \|b\|_2^2$ where $A \in \mathbb{R}^{n \times n}$ is a symmetric and positive definite matrix. Also assume $b \in \mathbb{R}^n$. Suppose you want to minimize the function $g(z)$ using the gradient descent algorithm.
- (a) Assuming the gradient descent iterates as $z_{k+1} = z_k + \alpha d_k$, derive the expression for d_k . [2 marks]
- (b) Derive a closed form expression for optimal stepsize for the $(k+1)^{th}$ iteration. You may define the optimal steps size as [2 marks]

$$\alpha = \underset{\alpha}{\operatorname{argmin}} g(z_k + \alpha d_k)$$

.

Q4 Answer the following questions with justifications.

- (1) Let $f(x) = ax^2 + bx + c$ where $a > 0$. We intend to find a local minimum of this function using gradient descent with hyperparameter λ . If x_n represents the value of the variable x after the n th step, is it possible to write $x_{n+k} = x_n P^k + Q$ where P and Q are constants? If so, find P and Q . Otherwise explain why you cannot write x_{n+k} in the form suggested. Does there exist a set of values of λ for which the algorithm is guaranteed to converge? If so find these values of λ . Give a mathematical justification for your answer. (5 Marks)
- (2) Let \mathbf{x} and \mathbf{z} be two $n \times 1$ vectors, for which a kernel function is defined as $K(\mathbf{x}, \mathbf{z}) = (\mathbf{x}^T \mathbf{z})^2 + 3(\mathbf{x}^T \mathbf{z} + 2)^2$. 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)