

Birla Institute of Technology and Science, Pilani

Work Integrated Learning Programmes Division

Cluster Programme - M.Tech. in AIML

I Semester 2023-24

Course Number	AIMLC ZC416	
Course Name	Mathematical Foundations for Machine Learning	
Nature of Exam	Open Book	# Pages 4
Weightage for grading	40%	# Questions 4
Duration	120 minutes	
Date of Exam	20/04/2024 AN	

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) Consider a function $f(x, y) = 3x^2 + 2y^2$. Assume that we use the gradient descent algorithm with momentum term to find the minimum of this function $f(x, y)$. Let the momentum/friction parameter used in this algorithm be referred to as β . Find the value of β if you are given the following information about the algorithm:

(i) Initial point of algorithm is $\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$

(ii) The iterates obtained after 3 iterations is given as $\begin{bmatrix} x_3 \\ y_3 \end{bmatrix} = \begin{bmatrix} -7.36 \\ 1.12 \end{bmatrix}$

- (iii) A fixed step size is used for all iterations and its value is $\alpha = 0.5$
(4 marks)

(B) Consider the positive definite matrix $\mathbf{A} = \begin{bmatrix} 4 & 2 & 6 \\ 2 & 17 & 7 \\ 6 & 7 & 35 \end{bmatrix}$

- (i) Calculate a lower triangular matrix \mathbf{L} such that $\mathbf{A} = \mathbf{L}\mathbf{L}^T$.

- (ii) Calculate all the eigenvalues of \mathbf{L} derived in (i).

(3 marks)

- (C) Consider a quadratic function $f(x, y) = x^2 + \beta y^2$ where $\beta \in \mathbb{R}$ is an unknown constant. Also assume that $\beta > 0$. Consider the problem of minimizing this function using gradient descent algorithm:

- (i) Derive a closed form expression (involving β) for the optimal step-size α for the first iteration of gradient descent if the initial point is

$$\begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

- (ii) If it is given to you that the optimal step-size $\alpha = 0.5$, derive the value of constant β using the formula derived in (i).

(3 marks)

Q2 Answer the following questions with justifications.

(A) Consider the data:

Positive class: $\{(7,0), (9,0)\}$ and Negative class: $\{(8,0), (10,0)\}$. Use the Kernel trick in SVM, find the suitable feature transformation $\phi(x)$. Find the decision boundary. (5 marks)

(B) For the data set $\mathbf{X} = [(\mathbf{4}, -\mathbf{3}), (\mathbf{0}, \mathbf{1})]$ and the feature transformation $\phi(x) = [x_1, x_2, ||x||]^T$, compute the Kernel matrix \mathbf{K} . (3 marks)

(C) For each of the following data samples, find the hinge loss. Also indicate which is the misclassified sample? (2 marks)

y	y'
0.5	1
1	-1

Q3 Answer the following questions with justifications.

(A) Consider a system of equations $\mathbf{A}\mathbf{x} = \mathbf{b}$ where \mathbf{A} is of order $m \times m$. Let $\mathbf{R}_i \in \mathbb{R}^m, i = 1, \dots, m$ be such that \mathbf{R}_i^T is the i th row of \mathbf{A} . It is given that each \mathbf{R}_i is orthogonal to \mathbf{x} for $i = 1, \dots, m$ with respect to the standard inner product.

(i) Find \mathbf{b} . (1 mark)

(ii) If $S = \{\mathbf{x} \in \mathbb{R}^m \mid \mathbf{R}_i \text{ is orthogonal to } \mathbf{x} \text{ for all } i = 1, \dots, m\}$ where orthogonality is with respect to the standard inner product. Then prove that S is a subspace of \mathbb{R}^m . (2 marks)

(iii) If $\text{rank}(\mathbf{A}) = m$, find dimension of S and S . (1 mark)

(B) Let $\sigma(z) = (1 + e^{-z})^{-1}$.

(i) Prove $\frac{d\sigma}{dz} = \sigma(z)(1 - \sigma(z))$. (1 mark)

(ii) If $f(x, y) = \alpha \ln\left(\frac{1}{\sigma(x + \beta y)}\right) + (1 - \alpha) \ln\left(\frac{1}{1 - \sigma(x + \beta y)}\right)$ then prove $f(x, y) = (1 - \alpha)(x + \beta y) - \ln(\sigma(x + \beta y))$. (2 marks)

(iii) Prove that $\frac{\partial f}{\partial x} = -\alpha + \sigma(x + \beta y)$ and $\frac{\partial f}{\partial y} = \beta(-\alpha + \sigma(x + \beta y))$. (2 marks)

(iv) Compute the Taylor's polynomial of degree 1 of f at $(0, 0)$. (1 mark)

Q4 Answer the following questions with justifications.

The figure below shows 4 points, representing some data in \mathbb{R}^2

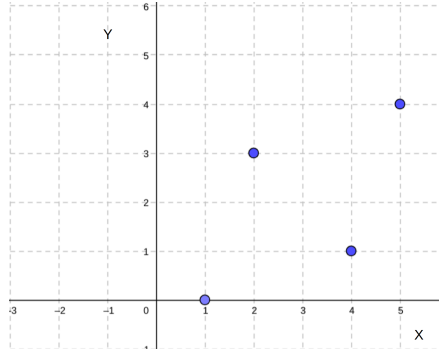


FIGURE 1. PCA

(A) Find the principal components for the given points.

NOTE: use

$$\text{cov}(X) = \frac{1}{N}(X - \mu)^T(X - \mu)$$

(5 marks)

(B) Find the components of the four points along their first principal component.

(2 marks)

(C) What is the percentage variance captured by the first principal component?

(1 mark)

(D) If the points are rotated anticlockwise by 90 degrees, what will the components (of the rotated points) along their first principal component be?

(2 marks)