2023-2024 SEM 1 AIML AB05-AA05 1-2023 Aimlczc 416 Mathematical Foundations FOR Machine Learning EC3 Regular 07-04-2024 AN

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
Weightage for grading	40%
Duration	120 minutes
Date of Exam	07/04/2024 AN

# 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) Consider a point in 2D space $\mathcal{P} = (4, 2)$ abd a line represented by the equation $f(x, y) = 0$. Using the method of Lagrange multipliers, derive the closest point on this line to the given point $\mathcal{P}$. You can assume that the closeness between two points is measured by square of euclidean distance. Derive the closest point to $\mathcal{P}$ when

i) $f(x, y) = x - 2y + 3$

ii) $f(x, y) = x + 2y + 5$

Also derive the distance to $\mathcal{P}$ in both cases. (4 marks)

(B) Consider the following matrix $\mathbf{C}$

$$\mathbf{C} = \begin{bmatrix} 6 & -33 & 19 \\ 6 & -8 & \frac{4}{3} \\ 9 & -33 & 16 \end{bmatrix}$$

i) Calculate the Trace of the matrix $\mathbf{C_1}$ where $\mathbf{C_1} = \mathbf{C^6}$

ii) Calculate the determinant of the matrix $\mathbf{C_2}$ where $\mathbf{C_2} = \mathbf{C^7}$

(3 marks)

(C) A professor gave his students three linearly independent vectors in R^n named $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$. He asked the students to construct three vectors $\mathbf{x}, \mathbf{y}$ and $\mathbf{z}$ defined as $\mathbf{x} = \mathbf{b} - \mathbf{c}$, $\mathbf{y} = \mathbf{a} + \mathbf{c}$ and $\mathbf{z} = \mathbf{a} - \mathbf{b}$. He asked students to consider a set of vectors named $\mathcal{H} = \{\mathbf{x}, \mathbf{y}, \mathbf{z}\}$. Prove or disprove that the set $\mathcal{H}$ is linearly independent. (3 marks)

Q2 Answer the following questions with justifications.

(A) Define the loss function

$$L(\boldsymbol{\beta}) = \frac{1}{2p}\|\boldsymbol{y} - \boldsymbol{\beta}\|^2 + \lambda\|\boldsymbol{W}\boldsymbol{\beta}\|^2$$

where $\boldsymbol{y} = [y_1, \cdots, y_p]^T, \boldsymbol{\beta} = [\beta_1, \cdots, \beta_p]^T \in \mathbb{R}^p, \lambda > 0, \boldsymbol{W} \in \mathbb{R}^{(p-2)\times p}$ and the norm is the Euclidean norm.

i) Show that we can write

$$L(\boldsymbol{\beta}) = \frac{1}{p}\sum_{j=1}^{p} L_j(\boldsymbol{\beta})$$

where $L_j(\boldsymbol{\beta})$ is independent of $y_i, i \neq j$. (1.5 marks)

ii) Also prove that

$$\nabla L_j(\boldsymbol{\beta}) = \left(\boldsymbol{v} + 2\lambda \boldsymbol{W}^T\boldsymbol{W}\boldsymbol{\beta}\right)^T, j = 1, \cdots p.$$

where $\boldsymbol{v} = [v_1, \cdots, v_p]^T$ is a p dimensional vector such that

$$v_i = \begin{cases} 0 & \text{when } i \neq j \\ -(y_j - \beta_j) & \text{when } i = j \end{cases}$$

(2 marks)

iii) To implement the gradient descent method for finding $\boldsymbol{\beta}$ that minimizes the loss function $L(\boldsymbol{\beta})$, find the gradient of $L(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$. (1.5 marks)

(B) A data analyst was interested in using Support Vector Machine (SVM) for binary classification. But the m dimensional data that he is interested in, is not linearly separable. So, one of his interns arrived at a feature map for the transformation of the data into a higher dimensional space. The feature map is defined as
$\boldsymbol{\phi}(\boldsymbol{x}) = [1, x_1, \cdots, x_m, x_1^3, \cdots, x_m^3, x_1x_2x_3, x_2x_3x_4, \cdots, x_{m-2}x_{m-1}x_m]^T$.
Find the dimension of the transformed data and find the corresponding kernel $\boldsymbol{K}(\boldsymbol{x}, \boldsymbol{y})$ that can be used in SVM. (1.5 marks)

(C) Let $f(x, y) = (x^2 + y^2)e^{-(x^2+y^2)}$ and A is the set of all possible critical points of f.

i) Find A. (1.5 marks)

ii) Prove or disprove that $[0, 0]^T$ is a point of minima of f. (2 marks)

Q3 Answer the following questions with justifications.

(A) The students were informed that the matrix $\boldsymbol{A}$ is a 6×6 matrix with real entries such that $\boldsymbol{A} = [\boldsymbol{C_1}, \boldsymbol{C_2}, \boldsymbol{C_3}, \boldsymbol{C_4}, \boldsymbol{C_5}, \boldsymbol{C_6}]$, where $\boldsymbol{C_i} \in \mathbb{R}^6$ and

$\boldsymbol{C_1} = \sum_{i=2}^{6} \boldsymbol{C_i}, \boldsymbol{C_2} = \sum_{i=3}^{6} \boldsymbol{C_i}$ and rank $(\boldsymbol{A}) = 4$.

i) A student claimed that in RREF($\boldsymbol{A}$), $\boldsymbol{C_1}$ is one of the non pivotal columns. If his claim is true, prove it, else provide a 6×6 matrix satisfying all the conditions given above but with RREF of that matrix having $\boldsymbol{C_1}$ as one of the pivotal columns. (3 marks)

ii) Another student claimed that $[1, -2, 0, 0, 0, 0]^T$ and $[-1, 0, 2, 2, 2, 2]^T$ satisfies homogeneous system of equations given by $\boldsymbol{AX} = \boldsymbol{0}$. Prove or disprove the student's claim. (2 marks)

(B) Consider a following dataset:

X_1	X_2	Y
2	2	Positive class
-2	-2	Positive class
2	-2	Negative class
-2	2	Negative class

i) Is the dataset linearly separable? (1 mark)

ii) Derive the appropriate Kernel Matrix for this problem. (1.5 marks)

iii) Derive the decision boundary using SVM. (2.5 marks)

Q4 Answer the following questions with justifications.

(A) Consider the following primal problem

$$
\begin{aligned}
&Min\ x_1^2 + x_2^2 - 4x_1 - 4x_2 \\
&\text{s.t. } x_1^2 \leq x_2 \\
&x_1 + x_2 \leq 2
\end{aligned}
$$

i) Find the dual of the above. (2 marks)

ii) Will dual and primal have same optimal function value? Justify. (2 marks)

(B) Teacher asked the students to conduct PCA on training dataset. He obtained eigenvalues of covariance matrix as 0.014, 0.016, 1, 3.5, 6.8, 12 and gave to the students.

i) He asked the students to find the dimension of each training sample. Students said its impossible to find given the insufficient information. Prove or disprove the claim made by the students. (1 mark)

ii) He further asked the students to find out the minimum number of principle components to retain after dimension reduction by PCA for the following cases:
a) If we want 95% of the variance to be retained.
b) If we want 99% of the variance to be retained.
Students said in both the cases we have to keep atleast 4 components. Prove or disprove the claim made by student. (3 marks)

iii) Teacher decided to keep the following two principle components:
$\begin{pmatrix} 0.115 \\ 0.106 \\ 0.118 \\ 0.082 \\ 0.013 \\ 0.023 \end{pmatrix}$ and $\begin{pmatrix} 0.205 \\ 0.215 \\ 0.315 \\ 0.428 \\ 0.238 \\ 0.034 \end{pmatrix}$.

Given a normalized training example $\begin{pmatrix} -3 \\ -2 \\ 2 \\ 1 \\ 2 \\ 4 \end{pmatrix}$, he asked students to find the result of PCA on this training example. Students said its not possible to apply. Prove or disprove students' claim. (2 marks)