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	120 minutes
Date of Exam	01/10/2022 (AN)

# Pages	3
# Questions	3

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 scientist works on a problem on $N > 10000$ data points on $D > 100$ dimensions, and obtains the top 10 principal components. Assume the given data is mean-centred. The data scientist's manager comes to him and tells him that due to a problem with the way in which the data was collected, every data point in D dimensions $(x_1, x_2, \ldots x_D)$ needs to be transformed to $(\alpha_1 x_1, \alpha_2 x_2, \ldots \alpha_D x_D)$ where $(\alpha_1, \alpha_2, \ldots, \alpha_D)$ are fixed constants. The data scientist believes that he can simply modify each discovered eigenvector $(b_1, b_2 \ldots b_D)$ to $(\alpha_1 b_1, \alpha_2 b_2, \ldots \alpha_D b_D)$ to get the principal directions for the modified problem. Is the data scientist correct in his belief? Give a mathematical justification for his belief. Otherwise explain why he is wrong. [5 Marks]

b Assume that a data scientist is originally given N points in D dimensions. In order to perform PCA the data scientist computes the covariance matrix. The data scientist is then informed that each given data point of the form $(x_1, x_2, \ldots x_d)$ needs to be transformed to $(\alpha_1 x_1, \alpha_2 x_2, \ldots, \alpha_D x_D)$ where $(\alpha_1, \alpha_2, \ldots, \alpha_D)$ are fixed constants. The data scientist thinks that he can compute the modified covariance matrix in only $O(D^2)$ time, given that he has the old covariance matrix with him. Is the data scientist justified in his belief? If so, demonstrate the method used by him. Otherwise, show why he is incorrect. [3 Marks]

c We are given three points and their associated classifications in the format $(x, y, \text{category})$ as follows: $(-1, 3, +), (1, 3, -), (-1, 1, -)$, and would like to find a separating hyperplane in the form $w^T \boldsymbol{x} + b = 0$ using SVM. Let $\alpha_i, 1 \leq i \leq 3$ be the Lagrangian multipliers for the given points. Set up the Lagrangian dual objective for this problem in terms of only the Lagrangian parameters as a polynomial in the fewest number of variables and the fewest number of terms. If possible, find the optimal separating hyperplane from this expression using the methods of calculus. Give adequate justifications. [8 Marks]

Q2 Answer the following questions with justifications.

a Assume that $x \in \mathbb{R}^n$. Consider the following functions :

i) $f_1(x) = \|x\|_1$. Prove or disprove that $f_1(x)$ is a convex function by using the properties of norms discussed in the course.

ii) Let $g(x)$ be a convex function and $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n$. Prove or disprove that $h(x) = g(Ax + b)$ is a convex function using the definition of convex functions discussed in the course. [4 marks]

b A professor was teaching singular value decomposition to students of an MTech programme. He asked them to consider a matrix $A \in \mathbb{R}^{n\times n}$.He also informed that its singular value decomposition is given by

$$A = U\Sigma V^T$$

where Σ is the matrix of singular values. It is also known that $\|\Sigma\|_F^2 = \gamma$ where $\|\cdot\|_F$ is the matrix Frobenius norm. Then professor defined a matrix B as $B = A^T A$ and a quantity

$$\alpha = B_{11} + B_{22} + ... + B_{nn}$$

A student named G1 claimed that $\alpha = \gamma$. Another student named G2 instead claimed that $\alpha = \sqrt{\gamma}$.

(a) Prove or disprove the claim made by G1

(b) Prove or disprove the claim made by G2

(c) Consider $C = AA^T$. Let

$$\beta = C_{11} + C_{22} + ... + C_{nn}$$

Prove or disprove that $\beta = \gamma^2$. [4 marks]

c Let $x \in \mathbb{R}^n$, $c \in \mathbb{R}^n$,$b \in \mathbb{R}^n$,$d \in \mathbb{R}$,$A \in \mathbb{R}^{n\times n}$,$A_1 \in \mathbb{R}^{n\times n}$ and $A_2 \in \mathbb{R}^{n\times n}$. Consider 2 functions

$$f(x) = \|Ax - b\|_2^2 + c^T x + d$$

and

$$g(x) = \|A_1^T A_1 x\|_2^2 + \|A_2^T x\|_2^2$$

i) Derive the gradient of $f(x)$ with respect to variable x.

ii) Derive the gradient of $g(x)$ with respect to variable x. [4 marks]

Q3 Answer the following questions with justifications.

a An analysis led to a matrix $\boldsymbol{A} = \begin{pmatrix} 1 & 0 & \rho \\ 0 & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$.

i) Find the conditions on ρ such that

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\boldsymbol{A}} = \mathbf{x}^T \boldsymbol{A} \mathbf{y}$$

is an innerproduct defined on $\mathbb{R}^3$. [3 marks]

ii) Is $\mathbf{x} = [1, 0, 0]^T$ perpendicular to $\mathbf{y} = [0, 1, 0]^T$ with respect to the inner product defined in (i)? Find all $\mathbf{z}$ perpendicular to both $\mathbf{x}$ and $\mathbf{y}$ with respect to the inner product defined in (i). [3 marks]

b A data scientist had arrived at a model $f(\mathbf{x}) = [\mathbf{x}^T \boldsymbol{Q}\mathbf{x}, \boldsymbol{b}^T\mathbf{x}]^T$ where $\boldsymbol{Q}$ is a 3×3 symmetric positive definite matrix and $\mathbf{x}, \boldsymbol{b} \in \mathbb{R}^3$.

i) Find the gradient $\nabla_{\mathbf{x}} f$. [1 mark]

ii) To simplify the computation, help him find the linear approximation of f about $(0, 0, 0)$. [2 marks]

c The Manager asked the data analyst to find a 2×2 matrix $\boldsymbol{M}$ with trace$(\boldsymbol{M}) = 0$ such that $(\|\boldsymbol{M} - \boldsymbol{A}\|_2)^2$ is minimum where $\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$. Help the data analyst in finding $\boldsymbol{M}$. [3 marks]