Indian Institute of Technology Madras (IIT Madras)

ID5055 Foundations of Machine learning

Mid-Term Exam Sample Question Paper Date: 27/09/2023 Time: 17:00–18:30 hrs

Instructions

1. The quiz is open class notes (your own handwritten), and class slides. The xerox copies of other handwritten notes and books are not allowed.
2. Write assumptions clearly to solve all the problems if it is not stated in the question.
3. All other rules and disciplinary actions of IIT Madras, Chennai for the Quizzes will be applied to this exam.

Questions

Answer the following Multiple Choice/Multiple Selection Questions. All the answers have to be supported by the explanation/calculations in the answer sheet(s).

Q1 Consider the data matrix collected from a manufacturing process involving eight variables. The following signular values (not ordered) are obtained after applying the Singular value decomposition (SVD): 1, 5.5, 30, 60, 2, 2, 100, 6. What is the minimum number of principal components that are required to explain around 98.5% variability in the data matrix?

(A) 3

(B) 4

(C) 5

(D) 6

Q2 For predicting price of a mobile brand (denoted by y), the data for three important features (denoted by x_1, x_2, and x_3) are collected along with the price. The data collected for 5 observations are given as follows

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \\ 5 \\ 3 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 3 \\ 3 \\ 2 \\ 3 \\ 5 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 5 \\ 7 \\ 4 \\ 13 \\ 11 \end{bmatrix} \tag{1}$$

The ML engineer considers the following three methods: (i) multiple linear regression, (ii) ridge regression, and (iii) least absolute shrinkage and selection operator (LASSO) regression approach. Which of the following option for the regression models can be applied to this dataset?

(A) (i), (ii)

(B) (i), (iii)

(C) (ii), (iii)

(D) (i), (ii), (iii)

Q3 A legal company has collected the following dataset for their cases to classify a court-case (y=civil (0) or crimincal (1)) based on two features x_1, x_2. The tuples (x_1, x_2, y) are given as follows: (2.5,1.5, 0), (3.5, 3.5, 1), (2, 3.5, 0), (2, 2, 0), (2.5, 2.5, 1), (2, 3, 1), (3.5, 3, 0), (2.5, 3, 0). A new point (3, 3, ?) has arrived. The DS wants to use k-nearest neighbor (kNN) algorithm to classify the new point using the Euclidean distance as a metric. Then, which of the following statement(s) is/are true?

(A) $k = 3$, the new point is classified as a civil case.

(B) $k = 5$, the new point is classified as a crimincal case

(C) $k = 4$, the new point is classified as a civil case

(D) $k = 6$, the new point is classified as a criminal case

Q4 Consider the following points: $\mathbf{o} = (1, 1)$, $\mathbf{p} = (1.5, 1.5)$, $\mathbf{q} = (1, 2.5)$, $\mathbf{r} = (\sqrt{2}, \sqrt{2})$, $\mathbf{s} = (3, 2.5)$, $\mathbf{t} = (4, 3.5)$. Consider the $\epsilon-$neighborhood represented by a circle with the radius =2, and the *MinPts* objects=3. Then, which of the following statements about these points is/are correct?

(A) **o**, **p**, **q**, and **r** are the core objects

(B) **o**, **p**, **q**, **r** and **s** are the core objects

(C) **p**, **q**, **r**, and **s** are density-reachable from **o**

(D) **p**, **q**, **r**, and **s** are density-reachable from **t**

Answer the following Numerical Answer type questions. All the answers have to be supported by appropriate calculations in the answer sheets(s).

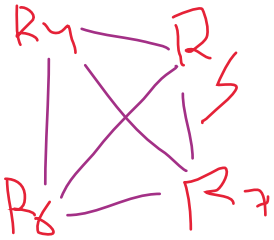
Q5 Consider the following datapoints collected for cases of fever in an institute: 1, 3, 2, 4, 4, 3, 2, 1, 5, 5. The following probability density function is assumed to be followed by the cases of fever.

$$f(x) = \frac{1}{\theta^2} x\, e^{-\frac{x}{\theta}} \tag{2}$$

What is the maximum likelihood estimate of θ?

Q6 Consider the following Laplacian matrix constructed from a similarity graph for a set of seven points.

$$\mathbf{L} = \begin{bmatrix} 2 & -1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ -1 & -1 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & -1 & -1 & -1 \\ 0 & 0 & 0 & -1 & 3 & -1 & -1 \\ 0 & 0 & 0 & -1 & -1 & 3 & -1 \\ 0 & 0 & 0 & -1 & -1 & -1 & 3 \end{bmatrix}$$



Find the number of connect components using the spectral clustering theory.

Q7 For a simple linear regression, the model equation between the independent variable (x_i) and dependent variable (y_i) is given by

$$y_i = \beta_1 x_i + \beta_0 + \epsilon_i \tag{3}$$

where $\epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$ is the model error in the ith observation with $\mathcal{N}$ being the normal distribution and σ_i^2 being the variance for the ith model error. For N observations of (x_i, y_i), derive the analytical expressions for β_0 and β_1.

Q8 Consider the following test error for different number of parameters as a tuple (Test error, number of parameters):

$$(121, 2), (100, 5), (80, 7), (79.5, 9), (90, 12), (95, 15)$$

Select the model with appropriate number of parameters and provide your reasons for the same.

Q9 For the case of orthonormal columns of $\mathbf{X}$ with N observations and p features (i.e., the p feature vectors are orthonormal), The estimates of β_j, $j = 1, \ldots, p$ for the ridge regression and lasso regression is given by

$$\beta_{j,ridge} = \frac{\hat{\beta}_j}{1 + \lambda}$$

$$\beta_{j,Lasso} = sign(\hat{\beta}_j)(|\hat{\beta}_j| - \lambda/2)_+$$

where $\hat{\beta}_j$ is the least-square (multiple linear regression) coefficient for the jth variable, $(z)_+ = max(z, 0)$.