INDIAN STATISTICAL INSTITUTE

(MS-QMS)_2025

Final-SEMESTER EXAMINATION (Pattern Recognition)

Duration: 180 minutes **Maximum Marks: 100**

Note: Answer any FIVE

Important Information:

Provide sufficient supporting descriptions and illustrations to your answers, and with an appropriate description of the used mathematical notations/symbols.

1.
   a. How does PCA partition the whole information space of the data set with variance? [7]
   b. How does variance quantify the loss incurred in using PCA? [2]
   c. Point out FOUR advantages and FOUR disadvantages of using PCA in data analysis in [4+2] in data analysis. Under which situation do PCA and LDA perform better than each other?
   d. Are the eigenvalues of a covariance matrix always real and positive definite? Justify your answer [5]

2.
a. What is the significance of logistic regression? and how is it different from linear regression and the OLS models? [3]
b. Derive the weight update equation for logistics regression. [7]
c. Formulate the working principle of logistic regression for a multiclass data set. Explain the one-versus-rest regression process with a small example dataset. [5]
d. List out at least TWO disadvantages of softmax regression, multiple regression and polynomial regression. [5]

3.
a. Provide the complete interpretation of the ROC curve. [5]
b. What is a perceptron? Provide a complete description of the perceptron, including its working principles, merits, and demerits. [8]
c. What are the changes required on the perceptron to use it as a linear regression? [2]
d. Describe the complete network architecture and working principle of an MLP neural network for the IRIS dataset. [5]

4.
a. Compare (merits and demerits) the KNN and K-means PR algorithms. [8]
b. Prepare a comparative table with merits and demerits for at least THREE activation functions (any). [6]
c. Prepare a comparative table listing the merits and demerits of THREE loss functions (any). [6]

5.
a. How is a Self-Organising Map different from MLP neural networks? [3]
b. Describe the working principle of SOM with a small example data set. [7]
c. What are the similarities and dissimilarities between ANN and CNN, in terms of architecture and functionalities [5]
d. Differentiate between vanishing gradient and exploding gradient. [5]
What problems do they create for the network, and how can their effect be minimised?

6.
a. Given an MLP neural network, list out SIX ways (approaches) to improve the MLP performance. [4]
b. List out THREE merits and TWO demerits of POOLING operation in CNN. [3+2]
c. Is standardization of the data desirable? Justify your answer [3]
d. Describe the DBSCAN clustering algorithm with its merits and demerits [8]

==========End of the Question paper==========