Reg. No															

B.Tech DEGREE EXAMINATION, MAY 2024

Fourth Semester

18AIC205J - NEURAL NETWORKS AND MACHINE LEARNING

(For the candidates admitted during the academic year 2018-2019 to 2021-2022)

Note:

i. **Part - A** should be answered in OMR sheet within first 40 minutes and OMR sheet should be handed over to hall invigilator at the end of 40th minute.

ii. **Part - B** and **Part - C** should be answered in answer booklet.

Time: 3 Hours **Max. Marks: 100**

PART - A (20 × 1 = 20 Marks)

Answer **all** Questions

		Marks	BL	CO
1.	In regression analysis, the variable that is being predicted is (A) Independent variable (B) Dependent variable (C) usually denoted by x (D) usually denoted by r	1	1	3
2.	In binary logistic regression (A) The dependent variable is continuous (B) The dependent variable is divided into two equal subcategories. (C) The dependent variable consists of two categories. (D) There is no dependent variable.	1	1	3
3.	One of the metrics used to assess a classification model is: (A) Confusion matrix (B) mean (C) standard deviation (D) variance	1	1	2
4.	The Euclidean Distance between the two data points A (1,3) and B (2,3) is (A) 1 (B) 2 (C) 3 (D) 4	1	1	4
5.	The difference between Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA) is (A) LDA is unsupervised learning algorithm while PCA is supervised (B) LDA assumes that the data is normally distributed while PCA does not make any assumptions about the data distribution (C) LDA is used for classification while PCA is used for dimensionality reduction (D) LDA is discriminative classifier while PCA is generative classifier	1	1	4
6.	What is the role of regularization in Regularized Linear Discriminant Analysis? (A) To prevent overfitting to the training data (B) To handle imbalanced datasets (C) To improve the accuracy of the classifier (D) To reduce the computational complexity of the algorithm	1	1	4
7.	The main goal of Principal Component Analysis (PCA) in machine learning is (A) To reduce the dimensionality of the dataset (B) To estimate the conditional probability of a class given a set of input features (C) To classify data into two or more classes based on input features (D) To perform regression	1	1	4

8.	In PCA, what is the impact of increasing the number of principal components? (A) It increases the amount of variance explained by the components (B) It reduces the dimensionality of the data (C) It has no effect on the amount of variance explained by the components (D) It decreases the amount of variance explained by the components	1	2	4
9.	Which of the following distance metrics is commonly used in K-Means clustering? (A) Euclidean distance (B) Manhattan distance (C) Minkowski distance (D) posterior probability	1	2	5
10.	Which of the following is a consequence of the curse of dimensionality? (A) The sparsity problem (B) The data scaling problem (C) The feature extraction problem (D) The bias-variance tradeoff	1	2	1
11.	What is the relationship between bias and variance? (A) Bias is the reciprocal of variance (B) bias is the multiplicative inverse of variance (C) bias is the additive inverse of variance (D) There is a tradeoff between bias and variance.	1	1	1
12.	How does a McCulloch-Pitts neuron model a logical OR operation? (A) By setting the threshold to 1 and applying a step function. (B) By setting the threshold to 0 and applying a step function. (C) By setting the threshold to 1 and applying a linear function. (D) By setting the threshold to 0 and applying a linear function.	1	2	1
13.	A neural network has an input layer with 5 neurons, a hidden layer with 10 neurons, and an output layer with 3 neurons. What is the total number of connections in the network? (A) 30 (B) 45 (C) 50 (D) 55	1	2	1
14.	An MLP is trained on a regression task to predict a single output variable. The loss function used during training is mean squared error (MSE). During testing, the predicted value for a particular input is 5.2, and the true value is 5.0. What is the MSE for this prediction? (A) 0.04 (B) 0.05 (C) 0.07 (D) 0.06	1	1	2
15.	What is the Kalman filter used for? (A) Image processing (B) Speech recognition (C) Object detection (D) State estimation	1	1	2
16.	Which of the following methods do we use to best fit the data in Logistic Regression? (A) Least Square Error (B) Maximum Likelihood (C) Jaccard distance (D) euclidean distance	1	1	3
17.	_____ assumes that the dependance of Y on X1, X2,..... X_n which is linear (A) Logistic Regression (B) Linear regression (C) Naïve Bayes (D) KNN Algorithm	1	1	3
18.	In a k-d tree, k originally meant? (A) Length of node (B) Size of tree (C) Weight of node (D) Number of dimensions	1	1	3
19.	Linear regression model based on mean squared error is most sensitive to (A) mean of input data (B) mean of output variables (C) number of features (D) outliers	1	1	4

		Marks	BL	CO
20.	How many coefficients is needed to estimate in a simple linear regression model? (A) 1 (B) 2 (C) 3 (D) 4	1	1	1

PART - B (5 × 4 = 20 Marks)

Answer **any 5** Questions

		Marks	BL	CO
21.	Imagine you're creating an application Z that can distinguish between photos of cats and dogs. You have a dataset with pictures of both cats and dogs. You want to use a neural network to classify the images. Describe the architecture of the neural network you'd use for this task.	4	4	1
22.	What is a Hidden Markov Model (HMM) and how does it differ from a traditional Markov Chain? Explain its components and its applications in real-world scenarios.	4	3	2
23.	Within the framework of binary logistic regression, the sigmoid function serves as a pivotal non-linear transformation, mapping the output of a linear function to the interval (0,1). Analyze the geometric attributes of its S-shaped curve and explain their significance in delineating a decision boundary for binary classification tasks.	4	3	3
24.	A researcher is attempting to employ a decision tree algorithm to predict student performance based on a dataset containing features like attendance rate, hours of study, and participation in extracurricular activities. Outline the potential risk of overfitting in decision trees and one method to mitigate it.	4	3	4
25.	In the context of machine learning and data analysis, explain the concept of the 'curse of dimensionality'. How does it impact the performance of algorithms and what are some common strategies employed to mitigate its effects	4	2	5
26.	Determine the machine learning method that employs the Bagging strategy. Additionally, outline the different phases of the method and provide a clear diagrammatic representation.	4	4	6
27.	Detail the methodologies encompassed in the Viterbi Algorithm's execution for determining the optimal sequence of hidden states. Illustrate its functioning with appropriate examples.	4	3	3

PART - C (5 × 12 = 60 Marks)

Answer **all** Questions

		Marks	BL	CO
28.	(a) Construct a single-layer perceptron and carry out two iterations. Initialize the perceptron with the weights w1=0.5 and w2=0, adopting a learning rate of α=0.2 and a bias value of θ=0.4. Employ the AND Boolean function for this exercise. Utilize the step activation function f(x) that produces outputs of either 0 or 1: it outputs 1 if f(x) is equal to or surpasses 0, and 0 otherwise. How does the perceptron perform after two iterations?	12	4	1
	(OR)			
	(b) Delve into the challenges associated with generalization in the realm of machine learning and explore potential solutions. Provide relevant examples and graphical representations to support your discussion.			
29.	(a) Elucidate the principles behind Hidden Markov Models (HMMs) and explore their practical applications in a real-world context. Suggest a dataset that aligns with the chosen context to showcase the utility of HMMs.	12	4	2
	(OR)			
	(b) Construct a Multi-Layer Perceptron implementing the backpropagation algorithm. Outline the architecture, detail the activation functions, and specify the learning parameters. Demonstrate its efficacy using a dataset of your choice. How would you validate its performance?			

30. (a) Given a dataset that lists the height, age, and weight (target) for individuals ID1 through ID10, you are faced with a challenge. An individual labeled as ID11 has a missing weight value. Use the Naïve Bayes Algorithm to deduce the weight of ID11 by leveraging the data from ID1 to ID10 and taking into account the height and age attributes of ID11. 12 4 3

(OR)

(b) Demonstrate the application of linear regression in fitting a curve given the subsequent data points:

x	y
1	3
2	5
3	7
4	9
5	11

31. (a) Determine the main component employing the PCA method for the provided dataset: { 2, 3, 4, 5, 6, 7 ; 1, 5, 3, 6, 7, 8 }. 12 3 4

(OR)

(b) Describe the fundamental mechanism of Support Vector Machines (SVMs). What are the optimal margin classifier function within SVMs, and explain the different regression variants associated with SVMs?

32. (a) Group the subsequent eight points, represented by their (x, y) coordinates, into three distinct clusters: A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9). Use the K-Means Algorithm with the initial cluster centers at A1(2, 10), A4(5, 8), and A7(1, 2), determine the positions of the three cluster centers after two iterations. 12 4 5

(OR)

(b) Given the sample dataset, there are four variables that help to predict the likelihood of an individual having heart disease. Based on this data, construct a Random Forest model to determine if an individual is at a risk of heart disease.

Blood Flow	Blocked Arteries	Chest Pain	Weight	Heart Disease
Abnormal	NO	NO	130	NO
Normal	YES	YES	195	YES
Normal	NO	YES	218	NO
Abnormal	YES	YES	180	YES

* * * * *