Reg. No															

B.Tech DEGREE EXAMINATION, DECEMBER 2023

Fourth Semester

18AIC205J - NEURAL NETWORKS AND MACHINE LEARNING

(For the candidates admitted during the academic year 2020 - 2021 & 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.	What is the name of the function in the following statement "A perceptron adds up all the weighted inputs it receives, and if it exceeds a certain value, it outputs a 1, otherwise, it just outputs a 0"? (A) Step function (B) Heaviside function (C) Logistic function (D) Perceptron function	1	1	1
2.	Backpropagation is a learning technique that adjusts weights in the neural network by propagating weight changes (A) Forward from source to sink (B) Backward from sink to hidden nodes to source (C) Backward from sink to source to hidden nodes (D) Forward from source to hidden nodes	1	1	1
3.	Which of the following is true for neural networks? (A) It is the same as logistic regression (B) It is the same as perceptron (C) It performs supervised learning (D) It cannot be used for classification	1	1	1
4.	When both inputs are 1 in the McCulloch-Pitts NOR gate model, what will be the output? (A) 1 (B) 0 (C) either 0 or 1 (D) z	1	2	1
5.	The positive sign of weights in the McCulloch-Pitts model is (A) Excitatory inputs (B) Inhibitory inputs (C) Neither excitatory nor inhibitory inputs (D) Bias	1	1	2
6.	You are building a binary classifier for recognizing cucumbers (y=1) vs. watermelons (y=0). Which one of these activation functions would you recommend using for the output layer? (A) ReLU (B) Leaky ReLU (C) Sigmoid (D) tanh	1	2	2
7.	The _____ is one way to quantify generalization error. (A) Bias (B) Variance (C) Bias-Variance Composition (D) Bias-Variance Decomposition	1	1	2
8.	What is the purpose of regularization in machine learning? (A) To prevent overfitting and improve generalization (B) To speed up the training process (C) To increase the accuracy of the model (D) To reduce the number of features in a model	1	1	2

9. Explain the concept of entropy in the context of decision trees. 1 1 3

(A) Entropy is a measure of impurity or disorder in a set of data
(B) Entropy is the number of leaf nodes in a decision tree.
(C) Entropy represents the number of features in a dataset.
(D) Entropy measures the height of a decision tree.

10. How can Naive Bayes handle continuous features? 1 1 3

(A) It ignores continuous features.
(B) It discretizes the continuous features into bins
(C) It splits the feature into ranges and computes the average over each range.
(D) It just assumes all features are categorical.

11. Suppose you are using RadialBasisFunction(RBF) kernel in SVM with a high Gamma value. What does this signify? 1 2 3

(A) The model would consider even far away points from the hyperplane for modeling
(B) The model would consider only the points close to the hyperplane for modeling
(C) The model would not be affected by the distance of points from the hyperplane for modeling
(D) The model will overfit.

12. Choosing _____values for k can be noisy and will have a higher influence on the result in KNN. 1 2 3

(A) Smaller
(B) Standard
(C) Larger
(D) binary

13. In PCA, the principal components are orthogonal. What does this mean? 1 1 4

(A) They have the same direction
(B) They are linearly independent
(C) They have zero variance
(D) They are negatively correlated

14. How is variance captured in PCA? 1 2 4

(A) By minimizing the mean
(B) By maximizing the eigenvalues
(C) By minimizing the eigenvalues
(D) By maximizing the covariance matrix

15. In LDA, how many linear discriminants can be derived if there are C classes? 1 2 4

(A) C
(B) C-1
(C) C+1
(D) 2C

16. In Kernel PCA, what is the role of the kernel function? 1 1 4

(A) To compute eigenvalues.
(B) To standardize the data.
(C) To map data into a higher-dimensional space
(D) To calculate the mean of the data.

17. What is the curse of dimensionality, and how does it impact clustering? 1 1 5

(A) The curse of dimensionality refers to the difficulty of visualizing high-dimensional data
(B) The curse of dimensionality causes clusters to be more compact.
(C) The curse of dimensionality does not affect clustering algorithms.
(D) The curse of dimensionality makes clustering more accurate.

18. Explain the concept of silhouette score in clustering. 1 1 5

(A) It measures the compactness of clusters.
(B) It evaluates the separation between clusters.
(C) It quantifies the goodness of fit of the clustering
(D) It is not related to clustering evaluation.

19. How does a Self-Organizing Map preserve the topology of the input space? 1 2 5

(A) By ignoring the input topology.
(B) By using a convolutional neural network.
(C) By organizing neurons in a grid that reflects the input space
(D) By converting input features into a higher-dimensional space.

20. Describe the dendrogram produced by agglomerative hierarchical clustering. 1 1 5

(A) A graphical representation of the data points.
(B) A tree-like structure showing the order of merging clusters
(C) A histogram of cluster sizes.
(D) A scatter plot of the clustered data.

PART - B (5 × 4 = 20 Marks) — Marks BL CO

Answer **any 5** Questions

21. Explain how the complexity of a machine learning model is related to overfitting and underfitting. Provide examples of situations where a model might be too complex or too simple. 4 2 1

22. Discuss the bias-variance tradeoff in the context of overfitting and underfitting. How does finding the right balance contribute to building a robust model? 4 2 1

23. Describe the architecture of a Multilayer Perceptron (MLP). How are the input, hidden, and output layers structured? What is the purpose of each layer? 4 2 2

24. Describe the logistic function (sigmoid function) used in logistic regression. How does it map real values to the range [0, 1]? 4 2 3

25. Discuss briefly different types of kernel functions commonly used in Kernel PCA, such as polynomial, radial basis function (RBF), and sigmoid kernels. 4 2 4

26. Explain the steps in the K-means clustering algorithm and apply the algorithm to cluster the following set of four objects into two clusters using K-means. The objects are A (3,5), B(4,5), C(1,3), and D(2,4). For initial cluster centers, consider the objects A and C. 4 3 5

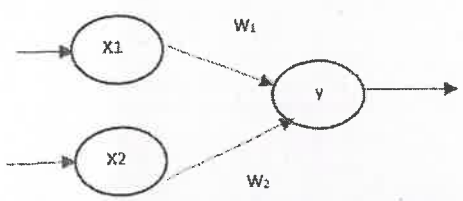
27. List and explain common performance metrics used in supervised learning, such as accuracy, precision, recall, and F1 score. 4 2 5

PART - C (5 × 12 = 60 Marks) — Marks BL CO

Answer **all** Questions

28. (a) Implement AND function using perceptron networks for bipolar inputs and targets. Assume your initial weights w_1=1.2, w_2=0.6 .bias=1 and learning rate=0.5. 12 3 1

X1	X2	y
1	1	1
1	-1	-1
-1	1	-1
-1	-1	-1



(OR)

(b) Draw the architecture of a single layer perceptron (SLP) and explain its operation. Mention its advantages and disadvantages.

29. (a) Describe the steps of the Backpropagation algorithm for training a neural network. Assume a simple feedforward neural network with one hidden layer and a mean squared error loss function. Provide details on each step of the algorithm, including forward pass, backward pass, and weight updates. 12 2 2

(OR)

(b) Explain the Viterbi algorithm in the context of Hidden Markov Models (HMMs). How does it work, and what is its primary application in HMMs?

30. (a) Describe different splitting criteria used in decision trees, such as Gini impurity and information gain. How does the choice of splitting criterion affect the construction of the decision tree? 12 2 3

(OR)

(b) Describe the least squares method used for estimating the coefficients in linear regression. How does it minimize the sum of squared differences between observed and predicted values?

31. (a) Consider a scenario where you have a dataset with a large number of features and a limited number of samples. Discuss how PCA and LDA can be used to address the challenges of the "curse of dimensionality. 12 3 4

(OR)

(b) Demonstrate a brief example illustrating the instantiation of the PCA class, its application to dataset fitting, and the subsequent transformation of data to reduced dimensionality. Please include sample code for clarity.

32. (a) Plot a dendogram using simple linkage (Agglomerative clustering) for the following data set. 12 3 5

Data Item	A	B	C	D	E
A	0	6	7	4	5
B	6	0	3	2	1
C	7	3	0	4	3
D	4	2	4	0	7
E	5	1	3	7	0

(OR)

(b) Construct Kohonen Self Organizing maps (KSOFM) with three input vectors X1(1,1,0), X2(1,0,0) and X3(1,0,1). Assume the number of clusters to be formed is 2 and the learning rate as 0.5. Initial weights C1: {0.2,0.4,0.6}, C2:{0.3,0.5,0.7}.

* * * * *