**B.Tech/ M.Tech (Integrated) DEGREE EXAMINATION, MAY 2024**
Fourth Semester

**21AIC202J – NEURAL NETWORKS AND MACHINE LEARNING**
*(For the candidates admitted from the academic year 2022-2023 onwards)*

**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: 75

**PART – A (20 × 1 = 20Marks)**
Answer **ALL** Questions

| | | Marks | BL | CO | PO |
|---|---|---|---|---|---|

1. In the context of version space learning, what is the role of a "Specific hypothesis"?     `Marks 1  BL 1  CO 1  PO 2`
   (A) It generalizes over positive examples
   (B) It narrows down the version space to a single hypothesis
   (C) It represents the entire space of positive hypothesis
   (D) It is only concerned with negative examples

2. How does the presence of noise in the data affect the well-posedness of a learning problem?     `Marks 1  BL 1  CO 1  PO 2`
   (A) It improves the stability of the learning problem
   (B) It makes the problem more well-posed by introducing randomness
   (C) It can degrade the well-posedness by introducing uncertainty
   (D) It has no impact on the well-posedness

3. In the candidate elimination algorithm, what does the hypothesis space represent?     `Marks 1  BL 1  CO 1  PO 2`
   (A) The set of all positive instances in the training data
   (B) The range of possible feature values
   (C) The set of all possible hypothesis consistent with observed examples
   (D) The total number of iterations during model training

4. How does the FIND-S algorithm handle negative examples in the training data?     `Marks 1  BL 2  CO 1  PO 2`
   (A) By adding them to the hypothesis set
   (B) By ignoring them during the algorithm
   (C) By removing inconsistent hypothesis form the set
   (D) By updating the hypothesis with each negative example

5. Which algorithm is commonly used for handling categorical features in decision trees?     `Marks 1  BL 1  CO 2  PO 2`
   (A) One-hot encoding
   (B) Label encoding
   (C) Ordinal encoding
   (D) Count encoding

6. What is the potential issue of decision trees that makes them susceptible to overfitting?    1 1 2 2
   - (A) They are sensitive to missing values in the data
   - (B) They are difficulty handling categorial features
   - (C) They can create overly complex trees that capture noise in the data
   - (D) They are limited to binary classification tasks

7. How does ID3 determine the attribute to split on at each node of the decision tree?    1 2 2 2
   - (A) By choosing the attribute with the highest information gain
   - (B) By selecting attributes randomly
   - (C) By considering attributes with the least number of missing values
   - (D) The choosing the attribute with the highest entropy

8. What happens if a node in the ID3 decision tree has no remaining attributes to split on but contains examples from multiple classes?    1 2 2 2
   - (A) The algorithm randomly selects an attribute
   - (B) The node becomes a lead and is labelled with the majority class
   - (C) The node is removed from the tree
   - (D) The algorithm selects a new attribute based on a predefined rule

9. What is the underlying assumption of the Naive Bayes classifier that makes it "Naive"?    1 1 3 2
   - (A) It assumes all features are independent
   - (B) It assumes all features are dependent
   - (C) It assumes features have equal importance
   - (D) It assumes continuous feature values

10. What is the purpose of the M-step in the EM algorithm?    1 2 3 2
   - (A) Maximizing the expectation of the complete data log-likelihood
   - (B) Maximizing the expectation of the incomplete data log-likelihood
   - (C) Minimizing the likelihood function
   - (D) Minimizing the sum of squared function

11. Which kernel is suitable for capturing interactions between features in a non-linear manner?    1 1 3 2
   - (A) Linear kernel
   - (B) Polynomial kernel
   - (C) Sigmoid kernel
   - (D) Laplacian kernel

12. In PCA, what is the relationship between the first principal component and the original features?    1 2 3 2
   - (A) It is a linear combination of all original features
   - (B) It represents the mean of the original features
   - (C) It is independent of the original features
   - (D) It is equal to the first original feature

13. What aspect of biological neural networks inspired the architecture of Artificial Neural Networks (ANNs)?    1 1 4 2
   - (A) Synaptic transmission
   - (B) Neuron activation functions
   - (C) Axon structure
   - (D) Dendritic spiking patterns

14. In a binary classification task, how does a perceptron make a decision?      1   2   4   2
     (A) By computing the weighted sum   (B) By using a random decision
         of inputs and applying an           boundary
         activation function
     (C) By computing the numbers of    (D) By comparing the mean of input
         input features                      features

15. How are weights assigned to data points in locally weighted regression?      1   2   4   2
     (A) Using a uniform weight for all   (B) Assigning weights based on the
         data points                         distance from the prediction point
     (C) Assigning random weights to    (D) Assigning weights based on the
         each data point                     value of the dependent variable

16. What is the cost function commonly used in logistic regression for model     1   1   4   2
    optimization?
     (A) Mean Squared Error (MSE)       (B) Cross-entropy loss (log loss)
     (C) Hinge loss                     (D) Huber loss

17. In a CNN, what is the role of the activation function in convolutional layers?   1   1   5   2
     (A) To increase the depth of the   (B) To compute the dot product
         network                            between weights and input
     (C) To introduce non-linearity in  (D) To normalize the input data
         feature maps

18. Which of the following is a variant of SGD that adapts the learning rate     1   1   5   2
    individually for each parameter?
     (A) Adaptive Gradient Descent      (B) Mini-batch stochastic gradient
         (ADAGRAD)                          descent
     (C) Batch gradient descent         (D) Learning rate annealing

19. Which activation function is characterized by a parameter 'OC' that controls     1   2   5   2
    the slope of the negative part, allowing it to take on any value between 0 and 1.
     (A) ReLU                           (B) Sigmoid
     (C) Tanh                           (D) Parametric ReLU (PReLU)

20. How does an LSTM cell differ from a traditional Recurrent Neural Network      1   2   5   2
    (RNN) cell?
     (A) LSTMs do not consider part     (B) LSTMs have fewer parameters
         information                        than RNNs
     (C) LSTMs do not use activation    (D) LSTMs have multiple gates to
         functions                          control the flow of information

## PART – B (5 × 8 = 40 Marks)    Marks   BL   CO   PO
### Answer ALL Questions

21. a. Given a weather dataset with the following examples in a binary classification     8   3   1   5
    problem.

| 1 | Sunny | Warm | Normal | String | Warm | Same   | Positive |
|---|-------|------|--------|--------|------|--------|----------|
| 2 | Sunny | Warm | High   | Strong | Warm | Same   | Positive |
| 3 | Rainy | Cold | High   | Strong | Warm | Change | Negative |
| 4 | Sunny | Warm | High   | Strong | Cool | Change | Positive |

Apply the FIND-S algorithm step by step and determine the final hypothesis.

**(OR)**

20MA4-21AIC202J

b. Imagine you are working on a project to develop a system that can predict whether a student will pass or fail a course based on historical data. Explain which type of learning (supervised, unsupervised or reinforcement) would be more suitable for this task. Justify your choice and outline the potential challenges.    8   4   1   5

22. a. Consider the following dataset that has three attributes 'A1', 'A2' and 'A3' and the target attribute 'classification'. Calculate the information gain of 'A1', 'A2' and 'A3' relative to this training instances. Also construct decision trees using ID3 algorithm.    8   4   2   5

| Instances | A1 | A2 | A3 | Classification |
|-----------|------|------|--------|----------------|
| 1 | True | Hot | High | No |
| 2 | True | Hot | High | No |
| 3 | False | Hot | High | Yes |
| 4 | False | Cool | Normal | Yes |
| 5 | False | Cool | Normal | Yes |
| 6 | True | Cool | High | No |
| 7 | True | Hot | High | No |
| 8 | True | Hot | Normal | Yes |
| 9 | False | Cool | Normal | Yes |
| 10 | False | Cool | High | Yes |

**(OR)**

b. Explain the concept of hypothesis space search in the context of decision tree learning. How does it relate to the process of finding an optimal tree for a given dataset?    8   2   2   5

23. a. Given the following instances, each instance has two attributes sepal length and sepal width. Compute the class label for test instance t1 (5.2, 3.1) when k = 5 using the KNN classifier.    8   3   3   5

| Sepal length | Sepal width | Species |
|--------------|-------------|------------|
| 5.3 | 3.7 | Setosa |
| 5.1 | 3.8 | Setosa |
| 7.2 | 3.0 | Virginica |
| 5.4 | 3.4 | Setosa |
| 5.1 | 3.3 | Setosa |
| 5.4 | 3.9 | Setosa |
| 7.4 | 2.8 | Virginica |
| 6.1 | 2.8 | Versicolor |
| 7.3 | 2.9 | Virginica |
| 6.0 | 2.7 | Versicolor |
| 5.8 | 2.8 | Virginica |
| 6.3 | 2.3 | Versicolor |
| 5.1 | 2.5 | Versicolor |
| 6.3 | 2.5 | Versicolor |
| 5.5 | 2.4 | Versicolor |

**(OR)**

b. How do SVM, handle non-linearly separable data? Discuss the concept of the kernel trick and its role in transforming data for better separability.    8   2   3   5

24. a. Draw the architecture of a single layer perception (SLP) and explain its operation. Mention its advantages and disadvantages.    8   2   4   5

**(OR)**

b. Provided below is a glucose prediction dataset with one independent variable, age (X) and one dependent variable glucose level (Y). Perform a simple linear regression to determine the equation of the best-fit line. Interpret the slope and intercept in the context of the problem. Also predict the glucose level of a patient when their age is 55.

| Subject | Age x | Glucose level y |
|---------|-------|-----------------|
| 1 | 43 | 99 |
| 2 | 21 | 65 |
| 3 | 25 | 79 |
| 4 | 42 | 75 |
| 5 | 57 | 87 |
| 6 | 59 | 81 |
| 7 | 55 | ? |

25. a. Provide a brief overview of the fundamental concepts behind Long Short-Term Memory (LSTM) networks. How do LSTMs address the vanishing gradient problem?
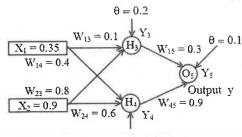
**(OR)**

b. Explain the purpose and functionality of convolutional layers in a convolutional neural network. How do convolutional layers help in feature extraction?

### PART – C (1 × 15 = 15 Marks)
### Answer **ANY ONE** Question

26. For the dataset represented below:
    (i)    Calculate the frequency tables for all these predictors
    (ii)   Calculate the likelihood tables for all three predictors
    (iii)   Using the Naive Bayes theorem identify if a given car is Red, SUV and domestic and will it get stolen or not.

| Color | Type | Origin | Stolen |
|-------|------|--------|--------|
| Red | Sports | Domestic | Yes |
| Red | Sports | Domestic | No |
| Red | Sports | Domestic | Yes |
| Yellow | Sports | Domestic | No |
| Yellow | Sports | Imported | Yes |
| Yellow | SUV | Imported | No |
| Yellow | SUV | Imported | Yes |
| Yellow | SUV | Domestic | No |
| Red | SUV | Imported | No |
| Red | Sports | Imported | Yes |

27. Consider the neural network below where the neurons have a sigmoid activation function. Perform a forward pass and a backward pass on the network. Assume that the actual output of Y is 0.5 and learning rate is 1, perform another forward pass. [θ-bias value].



\* \* \* \* \*