**D4ACS2302** **(2 Pages)** **Name.....**
**Reg.No.....**

**FOURTH SEMESTER M.Sc. DEGREE EXAMINATION, APRIL 2025**
**COMPUTER SCIENCE**
**FCSS4E04: ADVANCED MACHINE LEARNING**

**Time: 3 Hours** **Maximum Weightage: 30**

**Section A: Short answer. Answer any *four* questions. Each carries *two* weightage**

1. What does it mean for two vectors to be orthogonal? How do you compute the dot product to check for orthogonality? Check whether the two vectors are u=[2,3,-9] and v=[3,4,2] are orthogonal or not.
2. State and explain Bayes' Theorem. A diagnostic test is used to detect a rare disease that affects 1 in 1,000 people. The test has the following properties:

   - The probability of a positive result when the person has the disease (true positive rate, sensitivity) is 98%.
   - The probability of a negative result when the person does not have the disease (true negative rate, specificity) is 95%.

   A person tests positive for the disease. What is the probability that they actually have the disease, using Bayes' Theorem?
3. What is generalization in machine learning, and why is it important? How does overfitting affect the generalization ability of a model?
4. What are regression trees? How it is related to decision trees?
5. Describe the K-means clustering algorithm. How do you determine the optimal number of clusters? Discuss the advantages and limitations of K-means clustering.
6. What is model-based reinforcement learning? Discuss the differences between model-free and model-based approaches, and provide examples of each. How does a model-based approach help an agent learn from limited data?
7. How does the Universal Approximation Theorem address these limitations when using a Multi-Layer Perceptron (MLP)?

**(4 × 2 = 8 weightage)**

**Section B: Short Essay Question. Answer any *four* questions. Each carries *three* weightage**

8. What is the difference between a Bernoulli distribution and a Binomial distribution? Explain with an example.
9. Explain the chi-square distribution. What are some of its applications in hypothesis testing?
10. Assume you are working on a dataset where you need to predict a continuous target variable (e.g., house price) and a categorical target variable (e.g., loan approval status). How would you

use decision trees for both classification and regression tasks? Differentiate between both the trees.

11. What is the Bayesian approach to learning in machine learning? Explain how prior distributions and likelihood functions are combined to form the posterior distribution.
12. What is the concept of Dimensionality Reduction? Explain the significance with an example.
13. What is cross-validation, and how does it help in estimating the performance of machine learning models? Explain different types of cross-validation, such as k-fold cross-validation, leave-one-out cross-validation (LOO-CV), and stratified k-fold cross-validation.
14. Compare and contrast the Gradient Descent, Stochastic Gradient Descent (SGD), and Adam Optimizer

**(4 × 3 = 12 weightage)**

**Section C: Essay questions. Answer any *two* questions. Each carries *five* weightage**

15. Compare and contrast classification trees and regression trees. What are the main differences in terms of splitting criteria and applications?
16. Given the dataset which contains the students' interest in sitting the programming classes. The attribute "liked" is the label and other attributes are the decision based attributes. Compute the entropy of the dataset and also find out based on which attribute the first and second level decisions are to be made?

| Student | Prior Experience | Course | Time | Liked |
|---|---|---|---|---|
| 1 | Yes | Programming | Day | Yes |
| 2 | No | Programming | Day | No |
| 3 | Yes | History | Night | No |
| 4 | No | Programming | Night | Yes |
| 5 | Yes | English | Day | Yes |
| 6 | No | Programming | Day | No |
| 7 | Yes | Programming | Day | No |
| 8 | Yes | Mathematics | Night | Yes |
| 9 | Yes | Programming | Night | Yes |
| 10 | Yes | Programming | Night | No |

17. Explain the concept of spectral clustering. How does it use eigenvalues and eigenvectors of a similarity matrix to perform clustering? Discuss its advantages over traditional clustering algorithms like K-means.
18. Explain the use of statistical tests like the binomial test, approximate normal test, and t-test in evaluating machine learning algorithms. In which situations would you apply each of these tests to compare model performances?
19. Explain the concept of transfer learning in deep learning. How can a pre-trained network, such as a CNN trained on ImageNet, be fine-tuned for a different but related task?

**(2 × 5 = 10 weightage)**