Reg. No.

MANIPAL INSTITUTE OF TECHNOLOGY
MANIPAL

**II SEMSTER M. TECH. (COMPUTER SCIENCE & ENGINEERING) AND (COMPUTER SCEINCE & INFORMATION SECURITY)**
**ENDSEM SCHEME (MAY-2023)**

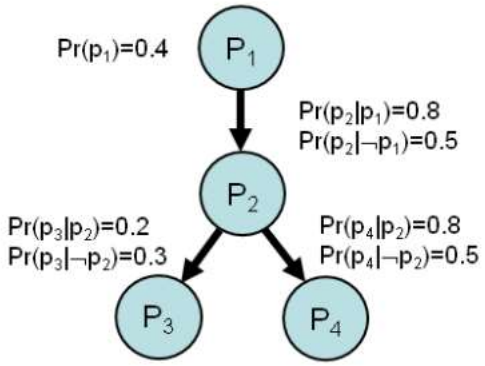
SUBJECT: ADVANCED MACHINE LEARNING-AML [CSE/CSIS 5256]

Time: 3 hrs. MAX. MARKS: 50M

**Instructions to Candidates:**

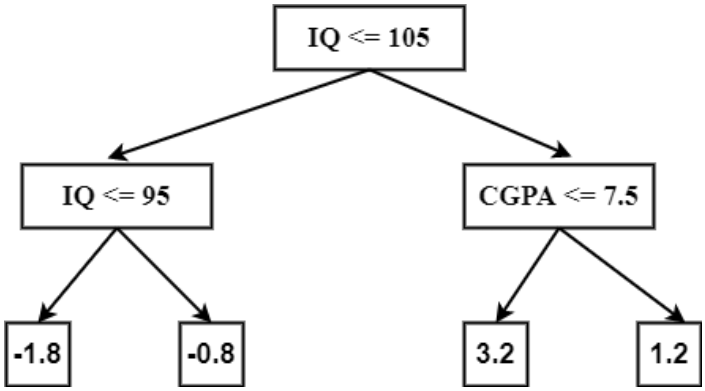
- Answer **ALL** the questions.
- Missing data may be suitably assumed.

| | | |
|---|---|---|
| Q1 a) | Suppose we have a dataset that consists of the heights and weights of a group of individuals, along with their gender (male or female). The dataset contains 10 individuals:<br><br>Height (inches) Weight (pounds) Gender<br>65 125 Female<br>71 175 Male<br>69 140 Female<br>64 130 Female<br>72 202 Male<br>73 200 Male<br>67 160 Female<br>70 172 Male<br>62 115 Female<br>71 190 Male<br>We want to make use of KNN to predict the gender of a new individual who is 68 inches tall and weighs 150 pounds. | 5M |
| Q1 b) | List some key differences between symbolic and connectionist approaches to learning, and which approach is better suited for concept learning versus machine learning? | 3M |
| Q1 c) | Illustrate some limitations of the Candidate Elimination algorithm? How can these limitations be addressed? | 2M |
| | | |
| Q2 a) | Suppose you have a dataset that contains information about customers of an online retailer. The dataset includes the age, income, and purchase history of each customer, as well as whether or not they are a frequent shopper (a binary variable). The retailer wants to use this data to predict which customers are likely to become frequent shoppers, so that they can target | 4M |

| | | |
|---|---|---|
| | them with special promotions. You are tasked with building a KNN classifier to solve this problem. | |
| Q2 b) | Identify the limitations of Branch and Bound MLKNN, and in what situations might it not be the best approach for KNN-based classification? | 3M |
| Q2 c) | Illustrate the procedure to measure the quality of a split in a decision tree using the Gini index, and how does it compare to other split criteria such as information gain? | 3M |
| | | |
| Q3 a) | Explain how do we compute the gradient of the error function with respect to the network weights during backpropagation? | 4M |
| Q3 b) | List the key steps involved in the backpropagation algorithm, and how do they contribute to the optimization process? | 3M |
| Q3 c) | Explain how does the RBF network compare to other machine learning algorithms, such as support vector machines and decision trees, in terms of accuracy and computational efficiency? | 3M |
| | | |
| Q4 a) | Given the Bayesian belief network in figure 1, calculate marginal and conditional probabilities P (¬p3), P (p2\|¬p3), P(p1\|p2, ¬p3) and P (p1\|¬p3, p4).<br>  <br><br>Fig. 1 | 5M |
| | | |
| Q4 b) | Suppose we have two features $\mathbf{x} = (x_1, x_2)$ and the two class-conditional densities, $p(x\|\omega = 1)$ and $p(x\|\omega = 2)$, are 2D Gaussian distributions centred at points (4, 11) and (10, 3) respectively with the same covariance matrix $\Sigma = 3I$ (with I is the identity matrix). Suppose the priors are $P(\omega = 1) = 0.6$ and $P(\omega = 2) = 0.4$.<br>i) Suppose we use a Bayes decision rule, write the two discriminant functions $g_1(x)$ and $g_2(x)$.<br>ii) Write the two discriminant functions $g_1(x)$ and $g_2(x)$ using Bayes decision rule | 3M |
| | | |
| Q4 c) | Compare k-medoid with k-means clustering technique with respect to robustness and computational complexity. | 2M |
| | | |
| | | |
| Q5 a) | Consider the dataset shown in Table 1 for which gradient boosting regressor is applied for prediction of salary based on {IQ, CGPA} as features . Here, | 4M |

| | | |
|---|---|---|
| | the decision tree regressor is used as base learning model where tree fitted for second base learner is shown in Figure 2. Using the gradient boosting algorithm, compute the pseudo-residuals(Res1,Res2) and predictions (Pred1,Pred2) for all instances in the dataset.<br><br>Fig. 2 | |

| Sl.No | IQ(x1) | CGPA(x2) | Salary(y) |
|---|---|---|---|
| 1. | 90 | 8 | 3 |
| 2. | 100 | 7 | 4 |
| 3. | 110 | 6 | 8 |
| 4. | 120 | 9 | 6 |
| 5. | 80 | 5 | 3 |




| | | |
|---|---|---|
| | | |
| Q5 b) | Justify a statement “Why does an ensemble model works”. Contrast between a bagging and boosting methods in ensemble learning. | 3M |
| | | |
| Q5 c) | Discuss the usefulness of following major tasks of clustering evaluation.<br><br>i) Assessing Clustering Tendency ii) Determine the Number of Clusters | 3M |
| | | |