Graphic Era
HILL UNIVERSITY
Established by an Act of the State Legislature of Uttarakhand (Adhiniyam Sankhya 12 of 2011)
University under section 2(f) of UGC Act, 1956

End Term (Odd) Semester Examination December 2024

Roll no. 236 11 45

Name of the Course and semester: B. Tech (3rd Semester)
Name of the Paper: Fundamentals of AI & ML
Paper Code: TCS-364
Time: 3 hour
100

Maximum Marks:

Note:

(i)   All the questions are compulsory.
(ii)   Answer any two sub questions from a, b and c in each main question.
(iii) Total marks for each question is 20 (twenty). Each sub-question carries 10 marks.

Q1.   (2X10=20 Marks)
a. How would you define "intelligent" in the context of an Intelligent Agent, describe its key characteristics, and analyze whether reflex actions, such as flinching from a hot stove, can be considered rational or intelligent? (CO1)
b. List two advantages and two disadvantages of using each of the following as a general-purpose AI knowledge representation scheme: (a)Natural Language (b)Standard Computer Science Databases (c)Frames (d)Semantic Networks (e)First Order Logic (f)If-Then Rule Based Systems

We previously saw that the Honda ASIMO is an impressive humanoid robot. What were the major knowledge representation issues that the developers needed to overcome to achieve their abilities? (CO2)
c. Identify **two applications** of AI in the field of healthcare and elaborate your thought on the following points:
a.   What does the application do?
b.   State what aspects of it make it an intelligent system. (CO3)

Q2.   (2X10=20 Marks)
a.   A small dataset contains the following points, where each point belongs to one of two classes (A or B):
Point (x, y)   Class
(2, 3)   A
(1, 1)   A
(4, 4)   B
(5, 2)   B
A new point P(3,2) needs to be classified using the **k-Nearest Neighbors (k-NN)** algorithm with k=3. Determine the class of P(3,2) based on majority voting, assume ties are resolved arbitrarily if needed. Explain one advantage and one limitation of the k-NN algorithm for this example. (CO3)
b. Compare uninformed search strategies with informed search strategies in solving AI problems. Provide examples of problems where informed search significantly outperforms uninformed search. (CO2)
c. What is 'Overfitting' and 'Underfitting' in Machine learning? Why it happens? How one can avoid these issues while performing machine learning? (CO4)

Q3.   (2X10=20 Marks)
a.   Russell & Norvig (2003) they list several definitions of AI from other textbooks and attempt to organize them along the two dimensions "human vs ideal" and thought vs. action" to give four categories:

Systems that think like humans	*Systems that think rationally*
Systems that act like humans	*Systems that act rationally*

How do you think Supervised systems fit in with the above classifications? What level of achievement is required of a supervised system before you would accept it as an example of AI? (CO1)
b. Elaborate your understandings on the following: (CO1, CO2)
•   The definition of an "agent" given in the lectures is quite broad. Can everything be described as an agent? What is an example of a non-agent? What about clocks – in what sense are they agents? Does the distinction between agents and non-agents really make any sense? (5)
•   Pick an Artificial Intelligence system that has appeared in the news or on TV recently and work out PAGE (i.e. Percepts, Actions, Goals, Environment) descriptions for it. How do these relate to their PEAS (i.e. Performance, Environment, Actuators, and Sensors) descriptions?
c. You are given a dataset with the following characteristics:
1.   The dataset contains 20,000 samples with 15 features.
2.   Some features are continuous, while others are categorical.
3.   The dataset is imbalanced, with 85% of the samples belonging to one class and 15% to the other.
4.   The decision boundary is expected to be nonlinear.
•   **Compare Logistic Regression, K-Nearest Neighbours (KNN), and Support Vector Machines (SVM)** as

End Term (Odd) Semester Examination December 2024

potential classification techniques for this dataset.

- Recommend the most appropriate classification technique for this dataset. Justify your choice with proper reasoning, considering the strengths and limitations of the selected model in this scenario. (CO3)

Q4. (2X10=20 Marks)

a. You are tasked with designing a simple machine learning pipeline for a classification problem where the objective is to predict whether a customer will buy a product (1 = "Yes", 0 = "No") based on the following features:

Age	Income (in $1000)	Gender (0 = Female, 1 = Male)	Marital Status (0 = Single, 1 = Married)
25	40	0	0
30	50	1	1
35	60	0	1
40	70	1	0
45	80	1	1

Data Preprocessing:
- How will you handle any missing values, if present.
- Convert categorical features (Gender and Marital Status) into numeric format using encoding techniques.

Model Selection and Evaluation:
- Choose an appropriate classification algorithm (e.g., Logistic Regression, Decision Tree, etc.) for this problem. Explain why you selected the model.
- How will you Split the data into a training and testing set.
- Evaluate the model's performance using accuracy, precision, and recall on the test set.
- Discuss how the performance metrics (accuracy, precision, recall) will help assess the model's usefulness in a real-world scenario. (CO6)

b. Define Accuracy, Precision, Recall, and F1-score and explain how they are calculated. Explain a scenario where **Accuracy** might be misleading and when **F1-score** should be preferred over Accuracy. (CO4)

c. Define a problem-solving framework for a taxi service that needs to optimize routes to pick up and drop off multiple passengers.
- Identify the state space, initial state, goal state, and actions.
- Propose a search strategy to solve the problem efficiently and justify your choice. (CO2)

Q5. (2X10=20 Marks)

a. A machine learning model was used to classify emails as Spam or Not Spam. The results of the classification are summarized in the following confusion matrix:

	Predicted: Spam	Predicted: Not Spam
Actual: Spam	80	20
Actual: Not Spam	10	90

Using the above confusion matrix:
1. Calculate the following metrics:
   - Accuracy
   - Precision for Spam
   - Recall for Spam
   - F1-Score for Spam
2. Interpret what the values of Precision and Recall indicate about the model's performance. (CO4)

b. The field of Artificial Intelligence emerged from several older disciplines. List the three that you think are most important and explain why. List the four sub-fields of AI that you think are most important. For each, give a representative goal or sub-goal or system that has been achieved so far. (CO1)

c. Consider the following axioms and answer the listed questions: (CO2)
1. All hounds howl at night.
2. Anyone who has any cats will not have any mice.
3. Light sleepers do not have anything which howls at night.
4. John has either a cat or a hound.

a. Convert the given axioms into predicate logic.
b. Convert them into CNF.
c. Prove the below statement using Resolution by refutation with resolution tree.
"If John is a light sleeper, then John does not have any mice."