②

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

Accelerated Semester

MID TERMEXAMINATION- 2026

B.Tech-VI Semester (CSE/IT)

COURSE CODE (CREDITS): 18B1WCI634 (2) MAX. MARKS: 50
COURSE NAME: MACHINE LEARNING
COURSE INSTRUCTORS: VANI SHARMA MAX. TIME: 2 Hours

Note: *(a) All questions are compulsory.*

(b) The candidate is allowed to make Suitable numeric assumptions wherever required for solving problems

(c) Calculator is allowed

<table>
<tr><th>Q.No</th><th>Question</th><th>CO</th><th>Marks</th></tr>
<tr><td>Q1</td><td>a) What is a ROC curve? Explain its importance in evaluating the performance of a classification model.

b) In a multiclass classification problem, the number of instances belonging to classes X, Y, and Z are 120, 100, and 80 respectively. The following confusion matrix is obtained after classification.
<table>
<tr><th>Actual/Predicted</th><th>X</th><th>Y</th><th>Z</th></tr>
<tr><td>X</td><td>90</td><td>20</td><td>10</td></tr>
<tr><td>Y</td><td>15</td><td>70</td><td>15</td></tr>
<tr><td>Z</td><td>5</td><td>10</td><td>65</td></tr>
</table>

Using the above confusion matrix, calculate the accuracy of the classifier and the precision and recall for each class.</td><td>[3]</td><td>[4+6]</td></tr>
<tr><td>Q2</td><td>a) Using the following dataset, predict the class for the record (Color-Red, Type-SUV, Origin=Domestic) using Naive Bayes algorithm.
<table>
<tr><th>Color</th><th>Type</th><th>Origin</th><th>Class</th></tr>
<tr><td>Red</td><td>Sports</td><td>Domestic</td><td>No</td></tr>
<tr><td>Red</td><td>Sports</td><td>Imported</td><td>No</td></tr>
<tr><td>Yellow</td><td>Sports</td><td>Domestic</td><td>Yes</td></tr>
<tr><td>Yellow</td><td>Sports</td><td>Imported</td><td>Yes</td></tr>
<tr><td>Yellow</td><td>SUV</td><td>Imported</td><td>Yes</td></tr>
<tr><td>Yellow</td><td>SUV</td><td>Domestic</td><td>No</td></tr>
<tr><td>Red</td><td>SUV</td><td>Imported</td><td>Yes</td></tr>
<tr><td>Red</td><td>SUV</td><td>Domestic</td><td>Yes</td></tr>
<tr><td>Yellow</td><td>Sports</td><td>Domestic</td><td>No</td></tr>
<tr><td>Red</td><td>Sports</td><td>Imported</td><td>Yes</td></tr>
</table>
</td><td>[3,1]</td><td>[7+3]</td></tr>
</table>

<table>
<tr><td></td><td>b) What is over-fitting and under-fitting in machine learning. Express in terms of bias and variance.</td><td></td><td></td></tr>
<tr><td>Q3</td><td>a) What do you mean by reinforcement learning?
b) What do you mean by inductive learning?
c) What do you mean by logistic regression? Derive its objective function.</td><td>[1,2]</td><td>[2+2+6]</td></tr>
<tr><td>Q4</td><td>Consider the following dataset that predicts whether students pass a Machine Learning course (Yes/No) based on their previous GPA (High, Medium, Low) and whether they studied or not. Assume ($log_2 3 = 1.6$).
a) Find the entropy $H(Passed)$.
b) Find the entropy of the attributes, i.e., $H(Passed|GPA)$ and $H(Passed|Studied)$.
c) Draw the complete decision tree that would be learned from this dataset.
Dataset
<table>
<tr><th>Instance</th><th>GPA</th><th>Studied</th><th>Passed</th></tr>
<tr><td>1</td><td>Low</td><td>False</td><td>No</td></tr>
<tr><td>2</td><td>Medium</td><td>False</td><td>Yes</td></tr>
<tr><td>3</td><td>Medium</td><td>True</td><td>No</td></tr>
<tr><td>4</td><td>High</td><td>True</td><td>Yes</td></tr>
<tr><td>5</td><td>High</td><td>False</td><td>Yes</td></tr>
<tr><td>6</td><td>Medium</td><td>True</td><td>Yes</td></tr>
</table></td><td>[4]</td><td>[2+3+5]</td></tr>
<tr><td>Q5</td><td>a) What is K in the K-Nearest Neighbors (KNN) algorithm? What happens to the decision boundary if the value of K keeps increasing? Also explain how the optimal value of K can be determined.
b) Differentiate between traditional programming and machine learning systems with respect to how they solve problems and learn from data.
c) Given the following training dataset, determine which sport Angelina will play if the value of K = 3 in the KNN algorithm and her age is 5. Use the Euclidean distance metric.
<table>
<tr><th>Name</th><th>Age</th><th>Gender</th><th>Class of Sports</th></tr>
<tr><td>Ajay</td><td>32</td><td>0</td><td>Football</td></tr>
<tr><td>Mark</td><td>40</td><td>0</td><td>Hockey</td></tr>
<tr><td>Sara</td><td>16</td><td>1</td><td>Cricket</td></tr>
<tr><td>Zaira</td><td>34</td><td>1</td><td>Cricket</td></tr>
<tr><td>Sachin</td><td>55</td><td>0</td><td>Hockey</td></tr>
<tr><td>Rahul</td><td>40</td><td>0</td><td>Cricket</td></tr>
<tr><td>Pooja</td><td>20</td><td>1</td><td>Hockey</td></tr>
<tr><td>Smith</td><td>15</td><td>0</td><td>Cricket</td></tr>
<tr><td>Laxmi</td><td>55</td><td>1</td><td>Football</td></tr>
<tr><td>Michael</td><td>15</td><td>0</td><td>Football</td></tr>
</table></td><td>[3,2]</td><td>[2+2+6]</td></tr>
</table>