

PART A (2 marks each)

1. **What is the difference between classification and regression ?**
2. **What is 'Naive' in a Naive Bayes ?**

3. Explain overfitting and underfitting ?
4. What do you mean by Associative Rule Mining ?
5. What is PCA ? When do you use it ?
6. Define SVD. Also write their applications.
7. Compare Reinforced Learning and Supervised Learning.
8. What is Markov Decision Process ?
9. Describe the structure of Artificial Neural Networks ?
10. Name and define techniques used to find similarities in the recommendation system.

PART B

1. Explain Decision Tree algorithm in detail. What are Entropy and Information gain in Decision tree algorithm ? [4+4]
2. What are Recommender Systems ? What is Collaborative filtering and Content Based filtering ? [4+4]
3. What is the motivation behind the Gaussian Mixture Model ? What is the relationship between k-means and PCA ? [4+4]
4. What is curse of dimensionality and how can unsupervised learning help with it ? Discuss various dimensionality reduction techniques. [4+4]
5. Explain the purpose of slack variables in SVM problem formulation. When would you use SVM over random forest and vice-versa ? [4+4]
6. Differentiate Q-learning and SARSA and when you would use each one explain with example. [4+4]

7. Transform the following data into decision tree based on Gini impurity criteria. [4+4]

Deadline ?	Is there a party ?	Lazy ?	Activity
Urgent	Yes	Yes	Party
Urgent	No	Yes	Study
Near	Yes	Yes	Party
None	Yes	No	Party
None	No	Yes	Pub
None	Yes	No	Party
Near	No	No	Study
Near	No	Yes	TV
Near	Yes	Yes	Party
Urgent	No	No	Study

PART C (15 marks each)

1. The following dataset will be used to learn a decision tree for predicting whether a person is happy (H) or sad (S) based on the color of their shoes, whether they wear a wig and number of masks they have.

Color	Wig	No. of Marks	(Output) Emotion
G	Y	2	S
G	N	2	S
G	N	2	S
B	N	2	S
B	N	2	H
R	N	2	H
R	N	2	H
R	N	2	H
R	Y	3	H

- (i) What is the H (Emotion | Wig=Y) ? [Show your calculations] [2]
- (ii) What is the H (Emotion | Masks=3) ? [Show your calculations] [2]

- (iii) Which attribute would the decision-tree building algorithm choose to use for the root of the tree (Assume no pruning). [Show your calculations] [2]
- (iv) Draw the full decision tree that would be learned for this data (assume no pruning). [2]
- (v) What would be the training set error for this dataset ? Express your answer as the percentage of records that would be misclassified. [Justify your answer with proper reasoning] [2]

The following two questions does not use the previous example dataset.

- (vi) Assuming that the output attribute can take two values. What is the maximum training set error that any dataset could possibly have ? [2]
- (vii) Construct an example dataset that achieves this maximum percentage training set error. (It must have two or fewer inputs and five or fewer records). [3]
2. Explain why precision or recall alone is not a sufficient measure to measure classifier performance in the case of binary classification. Discuss AUC-ROC curve used in classification problems.
3. Illustrate Markov Decision Problem in reinforcement learning with the help of an example.
4. Discuss Association rule mining in detail. A database has nine transactions. Let min-support = 22% and min- confidence = 70%. Find all frequent item sets by using Apriori Algorithm and generate association rule on this. TID is the transaction ID. (7+8)

TID	List of Items
T100	L1, L2, L5
T100	L2, L4
T100	L2, L3
T100	L1, L2, L4
T100	L1, L3
T100	L2, L3
T100	L1, L3
T100	L1, L2, L3, L5
T100	L1, L2, L3

5. Train a perceptron model to learn the following truth table :

X	Y	Output (Target)
1	1	0
1	0	0
0	1	1
0	0	1

Take the initial parameters as $w_0 = 0.03$, $w_1 = -0.04$, $w_2 = 0.04$, $n = 0.4$ and bias as -1 . Use threshold activation function [$g(h) = 1$, if $h > 0$; $g(h) = 0$, if $h \leq 0$] and sequential update strategy. Consider input samples in the same order as given in the above table. Show iteration steps and show in graph how the decision boundary changes.