125

CSE/PC/H/T/412/2024

BCSE 4th Year 1st Semester Examination 2024
Machine Learning (Hons.)

Time: 3 Hours **Total Marks: 100**

==========

[CO1] Q1 Compulsory – 15 marks

[CO2] Answer any two between Q2 - Q4 – 10+10 marks

[CO3] Answer any three between Q5 – Q8 – 10+10+10 marks

[CO4] Answer either Q9 or Q10 – 10 marks

[CO5] Answer either Q11 or Q12 – 10 marks

[CO6] Answer either Q13 or Q14 — 15 marks

==========

1. (a) What is Machine Learning? Compare between Supervised and Unsupervised learning. [3+4]

   (b) Under what circumstances Reinforcement Learning will be used? Explain with example. [5]

   (c) Write down the steps of developing a machine learning model. [3]

2. (a) How does train-test split affect the performance of a machine learning model? Discuss. [4]

   (b) What is a Hypothesis in Machine Learning? What is Hypothesis Space? [3+3]

3. (a) Define VC dimension. Explain using the concept of VC dimension: Why cannot a linear classifier solve an XOR problem? [2+3]

   (b) Consider a learning algorithm that aims to learn a concept class with a sample complexity function $m(\epsilon,\delta)=1/\epsilon \log 1/\delta$.

   If the desired error ϵ is set to 0.01 and the confidence level δ is set to 0.1, what is the minimum required sample size for the learning algorithm to PAC learn the concept class? Now, if the confidence level δ is increased to 0.05 while keeping the desired error ϵ at 0.01, what impact does this change have on the minimum required sample size for PAC learning? [5]

4. Describe the "Candidate Elimination" algorithm for generating general to specific hypotheses. State under what condition the candidate elimination algorithm will converge. [6+4]

5. (a) What is the limitation of Bayes classifier? How does Naïve Bayes Classification overcome this? [4]

(b) Suppose we have the following training data: [6]

Class	Feature 1	Feature 2
A	1	2
B	3	4
C	5	6

We also have a new point, x = (2, 3). Show the steps for the minimum distance classifier to classify this point.

6. (a) Consider the following training examples:

What is the information gain of f2 for the said dataset? Provide the equation for calculating the information gain as well as intermediate results. [6]

Instance	Class	f1	f2
1	+	T	T
2	+	T	T
3	+	T	F
4	+	F	F
5	-	F	T
6	-	F	T

(b) Suppose you have a dataset with a binary target variable, such as spam or not spam. You need to choose the best feature for a decision tree using information gain. In this scenario, how would the presence of outliers in a specific feature affect its information gain, and what factors should be considered when determining whether to include or exclude that feature in the decision tree? [4]

7. An automotive engineer collects data on car fuel efficiency and records the amount of fuel consumed in liters per 100 kilometers **(v)** along with the corresponding distance traveled before refueling in kilometers **(p)**. The data is coded as **v = (x - 5) / 10**, where **x** is the amount of fuel consumed.

The following statistics were found: S_{vv}=5.73, S_{pv}=1.688, S_{pp}=1.168, pmean=3.22, vmean=4.22.

Find the equation of the regression line of p on v in the form **p=a+bv**. [10]

8. (a) What is the significance of "support vectors" in SVM? Draw the decision surface for linear SVM considering a sample dataset. [6]

(b) What are kernels? Describe how kernels help in attaining the non-linearity in SVM. [4]

9. (a) Highlight a scenario where F1 score might not be the most appropriate metric. [4]

(b) Two pilot models, P1 and P2, are developed for a credit card fraud detection system. The dataset contains 600 transactions with a class distribution of fraud to non-fraud as 1:9. Model P1 correctly identified 480 instances with a class ratio of 1:5, while Model P2 correctly identified 550 instances with a class ratio of 1:7. Which model do you think is better, and why? Support your answer with proper performance metric computation. [6]

10. (a) How can a confusion matrix be used to derive various performance metrics. Explain with example. [4]

(b) Compare and contrast the Precision-Recall curve with the ROC curve. [3]

(c) How does a perfect model's ROC curve look like, and what does an AUC of 1 signify? [3]

11. (a) Explain the concept of layer-wise architecture in a neural network. How does the number of hidden layers and neurons in each layer impact the expressiveness of the network? [4]

(b) Illustrate the backpropagation algorithm and its connection to gradient descent. How does the chain rule play a crucial role in updating the weights during training? [6]

12. You are given the following dataset with two features and one output, Design a perceptron model with the threshold activation function h(x)=1 if x>=0.5; 0 otherwise. [10]

Starting with weight W= [0,0] and bias b = 0, use the delta learning rule to update the weight until convergence is achieved. **[do upto 2 iterations]**

A	B	Output
0	0	1
0	1	1
1	0	1
1	1	0

13. (a) What is the primary difference between partition based clustering algorithms and hierarchical clustering techniques? [2]

(b) Compare and contrast k-means and k-medoid algorithms. [5]

(c) Use k-means algorithm and Euclidean distance to cluster the following 8 examples into 3 clusters: [8]

A1=(2,9), A2=(2,5), A3=(8,5), A4=(5,8), A5=(6,5), A6=(6,4), A7=(1,1), A8=(5,9).

Assume the initial seeds (centers of each cluster) are A2, A4, and A6. Run the k-means algorithm for 1 iteration only. At the end of this iteration show the following:

i) The new clusters (i.e., the examples belonging to each cluster)

ii) The centers of the new clusters and

iii) Draw a 10 by 10 space with all the 8 points and show the clusters after the first iteration and the new centroids.

14. (a) How do the parameters of DBSCAN algorithm affect the results of clustering? Discuss with example. [4]

(b) Use the distance matrix given below and perform single linkage operation. Show your results by drawing a Dendrogram. The Dendrogram should clearly show the order in which the points are merged. [5]

	P1	P2	P3	P4	P5
P1	0.00	0.10	0.41	0.55	0.35
P2	0.10	0.00	0.64	0.47	0.98
P3	0.41	0.64	0.00	0.44	0.85
P4	0.55	0.47	0.44	0.00	0.76
P5	0.35	0.98	0.85	0.76	0.00

(c) Explain the concept of bagging and boosting. How does diversity among base models contribute to the effectiveness of ensemble learning? [6]
