**BCSE 4th Year 1st Semester Examination 2025**
**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 among 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 Supervised, Unsupervised, and Reinforcement learning. [3+5]

   (b) When do we use Semi-supervised Learning? [3]

   (c) Describe (with an example) learning as a search problem. [4]

2. (a) Explain using the concept of VC dimension: Why cannot a linear classifier solve an XOR problem? Can a rectangular plane separate the XOR problem? If yes, show the shattering of the XOR point. [6]

   (b) Briefly describe how the train-test ratio impacts generalization ability. [4]

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

   (b) Suppose you have a Hypothesis class of linear classifiers in $R^2$, and you are given that the VC dimension of this class is d=3. How many examples are needed to achieve $\epsilon = 0.1$ accuracy, and $\delta = 0.05$ confidence using a PAC learning algorithm? [4]

4. Suppose you are a doctor treating a patient who occasionally suffers from an allergic reaction. Your intuition tells you that the allergy is a direct consequence of certain foods the patient eat, place where it is eaten, time of day, day of week and the amount spent on food. So we gathered some data and the data is summarized in the table below. What is the hypothesis that fits this data using the Find-S algorithm? Show trace. [10]

| Number | Restaurant | Meal | Day | Cost | Reaction |
|---|---|---|---|---|---|
| 1 | Sam's | breakfast | Friday | cheap | yes |
| 2 | Hilton | lunch | Friday | expensive | no |
| 3 | Sam's | lunch | Saturday | cheap | yes |
| 4 | Denny's | breakfast | Sunday | cheap | no |
| 5 | Sam's | breakfast | Sunday | expensive | no |

5. Build a decision tree to classify the following patterns. Use the information gain criterion. Show all the calculations systematically. What Boolean function does the tree realize? [10]

| Density | Grain | Hardness | Class |
|---|---|---|---|
| Heavy | Small | Hard | Oak |
| Heavy | Large | Hard | Oak |
| Heavy | Small | Hard | Oak |
| Light | Large | Soft | Oak |
| Light | Large | Hard | Pine |
| Heavy | Small | Soft | Pine |
| Heavy | Large | Soft | Pine |
| Heavy | Small | Soft | Pine |

6. You are given the following task. [10]

The training set consists of the following points:
Points from Class 1: {(11, 11), (13, 11), (8, 10), (9, 9), (7, 7), (7, 5), (15, 3)}
Points from Class 2: {(7, 11), (15, 9), (15, 7), (13, 5), (14, 4), (9, 3), (11, 3)}

Draw the Scatter plot. Are the two classes linearly separable? Explain.
In the Scatter plot, also sketch the decision boundary that you would get using a nearest neighbor classifier (assume k=1 for kNN).
Classify the sample (6, 11) using the classifier thus obtained.

7. a) What is the limitation of Bayes classifier? How does Naïve Bayes Classification overcome this? [2+2]

b) Consider the following table. Find out the Regression equation of Y on X. [6]

| Hours of mixing (X) | Temperature of wood pulp (Y) |
|---|---|
| 2 | 21 |
| 4 | 27 |
| 6 | 29 |
| 8 | 64 |
| 10 | 86 |
| 12 | 92 |

8. (a) Write down the difference between hard and soft margin. [2]

(b) Justify (T/F) with reasoning: [4]

SVM performs well for noisy datasets as compared to multilayered perceptron.

(c) What are kernels? Explain the kernel trick. [2+2]

9. (a) Highlight a scenario where precision might not be the most appropriate metric. Which metric will you use then? [3]

(b) You want to solve a classification task. You first train your network on 20 samples. Training converges, but the training loss is very high. You then decide to train this network on 10,000 examples. Is your approach to fixing the problem correct? If yes, explain the most likely results of training with 10,000 examples. If not, give a solution to this problem. [3]

(c) You observe the following while fitting a linear regression to the data: As you increase the amount of training data, the test error decreases and the training error increases. The training error is quite low (almost what you expect it to), while the test error is much higher than the training error. What do you think is the main reason behind this behavior? Explain your answer. [4]

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

(b) Explain how a confusion matrix relates to ROC curve? [2]

(c) The table below shows a confusion matrix results obtained from two models for a two-class classification problem. Which of the models will perform better? Use the appropriate performance measures in order to provide evidence in support of your answer. [4]

| | Predicted + | Predicted - |
|---|---|---|
| True + | 10 | 14 |
| True - | 60 | 290 |

Model 1

| | Predicted + | Predicted - |
|---|---|---|
| True + | 10 | 2 |
| True - | 102 | 260 |

Model 2

11.(a) What is the role of weights and biases in a neural network? [3]

(b) Discuss on the vanishing gradient problem in connection to an MLP. [4]

(c) In what scenario would increasing the depth of a neural network be more beneficial than increasing its width? Discuss the trade-off. [3]

12. (a) Is learning on batches better than fitting the entire training dataset on a single go? Discuss. [4]

(b) Train (**do upto 2 iterations**) a single-layer perceptron to model the AND logic function for two binary inputs (x1,x2). [6]

Inputs: (x1,x2) = (0,0), (0,1), (1,0), (1,1).

Target Outputs: y=0,0,0,1

Start with:

Initial weights: w1=0.0, w2=0.0, bias b=0.0, Learning rate: $\eta$=0.1.

Tasks to do:

1. Manually train the perceptron using the perceptron learning rule.
2. Compute updated weights and bias.
3. Validate the perceptron's outputs after training.

13. (a) What is the primary difference between partition-based clustering algorithms and hierarchical clustering techniques? [3]

(b) State the advantages of k-medoid algorithm over k-means algorithm. [3]

(c) Describe CLARANS as a clustering algorithm. [5]

(d) A dataset contains four points (0,1), (1,0), (0,0), (1,1). If k-means clustering is used to group the data into two clusters, comment on the convergence of the cluster stability concerning cluster initialization. [4]

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

(b) Use Complete-linkage agglomerative clustering to cluster the following 8 points: A1=(3,7), A2=(3,5), A3=(9,4), A4=(5,9), A5=(7,6), A6=(7,4), A7=(2,2), A8=(4,10). Show the dendrograms. Show all calculations. [6]

(c) Explain the difference between under-fitting and over-fitting. How can you detect each one? Write down various ways to mitigate under-fitting and over-fitting. [5]