MACHINE LEARNING
(CSBS 3232)

Time Allotted : 2½ hrs **Full Marks : 60**

Figures out of the right margin indicate full marks.

Candidates are required to answer Group A and any 4 (four) from Group B to E, taking one from each group.

Candidates are required to give answer in their own words as far as practicable.

Group – A

1. Answer any twelve: **12 × 1 = 12**

Choose the correct alternative for the following

(i) Which of the following is NOT typically included in input representation?
(a) Features (b) Noise
(c) Training Data (d) Labels

(ii) Which of the following metrics is most useful to detect overfitting in a model?
(a) High training accuracy and low test accuracy
(b) Low training accuracy and high test accuracy
(c) High training accuracy and high test accuracy
(d) Low training accuracy and low test accuracy

(iii) The VC Dimension of a perceptron in d-dimensional space is:
(a) 2d (b) d – 1
(c) d + 1 (d) Infinite

(iv) In a neural network, if the learning rate is 0.01, the gradient is -0.5, and the previous weight was 0.8, what is the updated weight using gradient descent?
(a) 0.794 (b) 0.815
(c) 0.648 (d) 0.8.5

(v) What is the purpose of likelihood in regression models?
(a) To measure the probability of data given a model
(b) To eliminate outliers
(c) To normalize data
(d) To reduce computation time

(vi) Recall is also known as
(a) Accuracy (b) Specificity
(c) False positive rate (d) Sensitivity

(vii) Which component in a CNN is responsible for feature extraction?
(a) Softmax layer (b) Fully connected layer
(c) Convolutional layer (d) Pooling layer

(viii) What is the purpose of input representation in machine learning?
(a) To store raw data without modification
(b) To increase computation time
(c) To transform data into a format suitable for learning
(d) To delete unnecessary data

(ix) Which technique is most effective for reducing high variance in a model?
(a) Using a more complex model
(b) Reducing the number of training samples
(c) Increasing the number of parameters
(d) Adding more training data

(x) A neural network has 3 input neurons, 2 hidden layers with 4 neurons each, and 2 output neurons. How many total connections (weights) are there, assuming fully connected layers?
(a) 26 (b) 36 (c) 34 (d) 32

Fill in the blanks with the correct word

(xi) A key limitation of linear regression is that it cannot model _____ relationships between input and output variables.

(xii) For a hypothesis class that can shatter 6 points but not 7 points, the VC dimension is _____.

(xiii) If a model misclassifies 30 out of 150 test samples, the accuracy of the model is ____%.

(xiv) A commonly used technique to prevent overfitting by adding a penalty term to the loss function is called _____.

(xv) The points that lie closest to the decision boundary in an SVM and influence its position are known as _____.

Group - B

2. (a) Compare between classification and regression algorithms. *[(CO1)(Understand/LOCQ)]*
(b) State the main differences between supervised learning and reinforcement learning. *[(CO1)(Understand/LOCQ)]*
(c) What is regularization? Explain L1 and L2-regularization in the formulation of objective functions in machine learning algorithms. *[(CO2)(Understand/LOCQ)]*

3 + 3 + (2 + 4) = 12

3. (a) Using gradient descent for linear regression, assume the cost function is given by:

$$J(w) = \frac{1}{m}\sum_{i=1}^{m}(y_i - wx_i)^2$$

Given an initial weight w=0.5, learning rate α=0.1, and data points (x_1, y_1)= (1, 2) and (x_2, y_2)=(2, 2.8), perform two iterations of gradient descent and compute the updated weight. *[(CO2, CO6)(Apply/IOCQ)]*
(b) For a logistic regression model with two input features x_1, x_2, the weight vector is given as w = [2, −1] and bias b=0.5. If a new data point (x_1, x_2) = (3,2) is given, compute the output of the logistic function and determine whether the model classifies it as class 0 or class 1. *[(CO2, CO6)(Apply/IOCQ)]*

6 + 6 = 12

Group – C

4. (a) "For unbalanced dataset, accuracy is not a good metric for classification" – justify with an example. *[(CO3)(Analyse/HOCQ)]*

   (b) There are 550 samples of men or women in a given dataset. Out of 60% women, 75% are classified correctly, and the remaining women are classified as men by a machine learning algorithm. On the other hand, in the men category, 85% are classified as men, and the remaining are classified as women.
   i. Derive the confusion matrix.
   ii. Calculate accuracy, precision, recall and F1 score. *[(CO3, CO6)(Apply/IOCQ)]*

4 + (3 + 2 + 1 + 1 + 1) = 12

5. (a) A machine learning model is trained with increasing dataset sizes. The training error and validation error are recorded:

Training Size	Training Error	Validation Error
50	0.08	0.2
100	0.06	0.18
200	0.05	0.12
500	0.04	0.07
1000	0.03	0.06

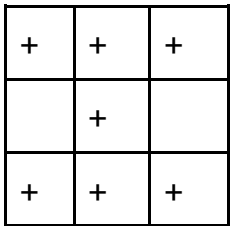
   Identify if the model is overfitting or underfitting. Predict what would happen if more training data is added. *[(CO3, CO6)(Apply/IOCQ)]*

   (b) Explain the bias-variance tradeoff with an example. How does increasing the complexity of a model affect bias and variance? *[(CO3)(Understand/LOCQ)]*

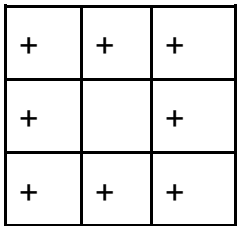
(3 + 3) + (3 + 3) = 12

Group - D

6. Using the Hebb rule, find the weights required to perform the following classifications of the give it input patterns shown below. The pattern is shown as 3 x 3 matrix form in the squares. The "+" symbols represent the value" 1" and empty squares indicate "-1." Consider "I" belongs to the members of class and "0" does not belong to the members of class.



I

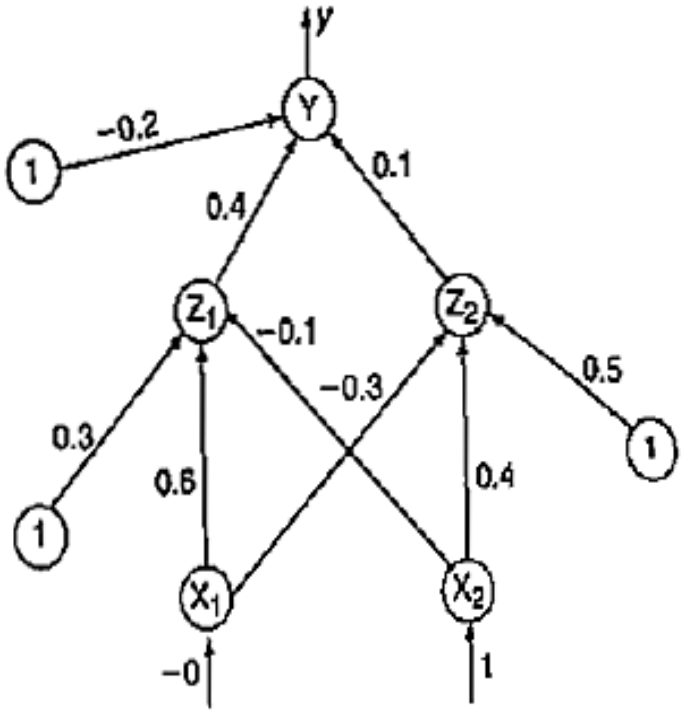


O

[(CO4, CO6)(Apply/IOCQ)]

12

7. Using a back-propagation network, find the new weights shown in Figure below. It is presented with the input pattern [0, 1] and the target output is 1. Use a learning rate = 0.25 and binary sigmoidal activation function.



[(CO4, CO6)(Apply/IOCQ)]

12

Group - E

8. Find the equation of a decision boundary using SVM to classify the following data points:

Data point		Class level
D1	D2	
2	3	X
2	4	X
3	1	Y

[(CO5, CO6)(Apply/IOCQ)]

12

9. Consider a two-dimensional dataset with two classes:
Class 1: (0,2),(1,3),(2,2)
Class 2: (2,0),(3,−1),(1,−2)
i. Identify the support vectors and compute the margin width.
ii. Predict the class of the unknown point (1.5, 1.5).

[(CO5, CO6)(Apply/IOCQ)]

(6 + 6) = 12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	18.75	77.08	4.17