FUNDAMENTALS OF MACHINE LEARNING
(CSEN 3203)

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 a predictive model?
(a) Clustering (b) Regression
(c) Summarization (d) Associationrule.

(ii) Data used to optimize the parameter settings of a supervised learning model
(a) training (b) test
(c) verification (d) validation

(iii) Statement 1: The error of a hypothesis measured over a training set provides a pessimistically biased estimate of true error of the hypothesis
Statement 2: Gradient descent has the problem of falling into local minima
(a) Only Statement 1 is true (b) Only Statement 2 is true
(c) Both are true (d) Both are false.

(iv) It is given that the break point of a hypothesis h(x) is n. This means
(a) h(x) cannot shatter any set of n data points
(b) h(x) cannot shatter any n+1 data points
(c) h(x) can shatter at least one training data set of n points
(d) All of the above

(v) The growth function h(N) for positive intervals (h(X) = 1 when $a \le X \le b$ and h(X) = -1 otherwise) is
(a) N+1 (b) ½ N(N+1) (c) 2^N (d) ½ N(N+1)+1

(vi) The back-propagation algorithm learns a globally optimal neural network with hidden layers
(a) AlwaysTrue (b) Always False
(c) Mostly true (d) Mostly False

(vii) Which of the following methods do we use to best fit the data in logistic regression?
(a) Least Square Error (b) Maximum Likelihood
(c) Jaccarddistance (d) both (a)and (b)

(viii) Which of the following gives non-linearity to a neural network?
(a) Convolution operator (weighted sum of the inputs)
(b) Stochastic gradient descent
(c) Sigmoid activation
(d) Non-zero bias.

(ix) A hard margin of SVM in a linearly separable data set indicates
(a) that the SVM allows very low error in classification
(b) that the SVM allows high amount of error in classification
(c) that all training data points are equally important to the SVM
(d) that the SVM guarantees that the in-sample error $E_{in}=0$

(x) Which regularization technique adds a penalty term to the cost function that is proportional to the sum of the squares of the model's weights?
(a) L1 regularization (b) Dropout regularization
(c) Ridge regularization (d) Elastic Net regularization.

Fill in the blanks with the correct word

(xi) When a model performs well on training data (the data on which the algorithm was trained) but does not perform well on test data (new or unseen data), we say that the model is _____.

(xii) In an MLP, the number of nodes in the input layer is 10 (including bias node) and in the hidden layer is 5 (including bias node). The maximum number of connections from the input layer to the hidden layer is _____.

(xiii) Classification can be categorised as a problem of _____ learning.

(xiv) Support Vector Machines (SVM) aim to maximize the _____ between the class boundaries.

(xv) In-sample error refers to the error calculated on the data used for _____.

Group – B

2. (a) Explain mathematically how the weight and bias expressions are determined for a Linear Regression Model utilizing the Least Squares Method on a single dependent variable. *[(CO2)(Understand/LOCQ)]*

(b) Apply the Least Squares Method on the given dataset to predict the output value for the test value of "X=4.5"

Input – X	1	2	3	4	5
Output – Y	3	7	8	12	15

[(CO6)(Apply/LOCQ)]

8 + 4 = 12

3. (a) Define Precision and Recall in the context of classification. *[(CO1)(Understand/LOCQ)]*

(b) Consider the following confusion matrix for a two-class data set on which classification has been done:

	Actual class 1	Actual class 2
Placed in class 1	34	26
Placed in class 2	36	44

Evaluate the overall accuracy of this classification results. Also calculate the Precision and Recall values. *[(CO4)(Analyze/IOCQ)]*

(c) Why do we need an iterative approach for solving linear regression? *[(CO2)(Understand/LOCQ)]*

(d) Explain the intuition behind Gradient Descent Algorithm. Compare Batch Gradient Descent with the stochastic one. *[(CO3)(Understand/LOCQ)]*

2 + (2 + 2) + 2 + (2 + 2) = 12

Group - C

4. (a) What are the in-sample and out-of-sample errors in the context of classification? *[(CO1)(Remember/LOCQ)]*

(b) Define Hoeffding's inequality in the context of feasibility of learning. *[(CO5)(Analyze/HOCQ)]*

(c) Briefly explain error and noise with example. *[(CO3)(Understand/LOCQ)]*

(d) Define VC Dimension. *[(CO1)(Remember/LOCQ)]*

2 + 2 + 6 + 2 = 12

5. (a) Define dichotomy, growth function and break point. *[(CO1)(Remember/LOCQ)]*

(b) Calculate the growth function and break point for any N number of data points for

(i) Positive rays: H consists of all hypotheses
$h(x) = +1$ when $x \geq a$ and $h(x) = -1$ when $x < a$;

(ii) Convex sets: H consists of all hypotheses in two dimensions
$h : R2 \rightarrow \{ - 1 , + 1\}$ that are positive inside some convex set and negative elsewhere. Show all the steps to justify your answers. *[(CO6)(Apply/IOCQ)]*

6 + 6 = 12

Group – D

6. (a) Describe the Perceptron Learning Algorithm (PLA) and provide a brief explanation of how does it work. *[(CO2)(Understand/LOCQ)]*

(b) Analyse the proof of convergence of the Perceptron Learning Algorithm (PLA) *[(CO4)(Analyse/HOCQ)]*

5 + 7 = 12

7. (a) Given the input matrix and the kernel, perform convolution with stride being 1.

1	0	0
0	0	1
1	1	0
Kernel Matrix		

1	0	1	1	0
0	0	0	1	1
1	0	0	0	1
0	1	1	1	0
1	1	0	1	0
Input Matrix				

[(CO5)(Apply/IOCQ)]

(b) Apply max-pooling and sum-pooling to the results from the above convolutions.

[(CO4)(Apply/IOCQ)]

4 + (4 + 4) = 12

Group – E

8. Consider a dataset with features (X_1, X_2) and target y with three data points as follows:

X_1	X_2	y
4	1	1.2
16	2	4.7
9	1	2.8

Let us set up a linear regression model $y = \theta_1 X_1 + \theta_2 X_2$ with no intercept term. Use a regularization hyper parameter 1 and learning rate of 0.01. From an initial guess of (θ_1, θ_2) = (2, 2) perform one step of batch gradient descent for both Ridge objective and Lasso objective. *[(CO6)(Apply/HOCQ)]*

(6 + 6) = 12

9. (a) What is the intuition of a large margin classifier? *[(CO1)(Remember/LOCQ)]*
   (b) What is a kernel in SVM? Why do we use kernels in SVM? Can we apply kernel trick in logistic regression? *[(CO1)(Remember/LOCQ)]*
   (c) Justify whether SVM is sensitive to the feature scaling. *[(CO3)(Evaluate/HOCQ)]*
   (d) Explain soft margin and hard margin Support Vector Machine. *[(CO2)(Understand/LOCQ)]*

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

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	52.1	22.9	25

Course Outcome (CO):

After the completion of the course students will be able to

CSEN3203.1. Learn and Understand the basics of Machine Learning Approaches and paradigm.
CSEN3203.2 Understand and describe various Machine Learning algorithm.
CSEN3203.3 Understand complexity of Machine Learning Algorithms and their limitations.
CSEN3203.4 Mathematically analyse various machine learning approaches and paradigm.
CSEN3203.5 Analyse various machine learning techniques to get an insight of when to apply a particular machine learning approach.
CSEN3203.6 Apply common machine learning algorithms in practice and implementing their own using real world data.

**LOCQ: Lower Order Cognitive Question; IOCQ: Intermediate Order Cognitive Question; HOCQ: Higher Order Cognitive Question.*