

**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) Everyday your email account is flooded with many emails. Problem 1: You want to know how many spam emails will be coming in the next one month. Problem 2: You want to know whether the upcoming email is spam or not. Which one the following is correct?
 - (a) Both are Classification Problems
 - (b) Problem 1: Classification, Problem 2: Regression
 - (c) Problem 1: Regression, Problem 2: Classification
 - (d) Both are Regression Problems
- (ii) In a linear regression problem, $h(x)$ is the predicted value of the target variable, y is the actual value of the target variable, m is the number of training examples. What do we try to minimize?
 - (a) $(h(x)-y)/m$
 - (b) $(h(x)-y)/(2*m)$
 - (c) $(h(x)-y)^2/(2*m)$
 - (d) $(y-h(x))$
- (iii) It is given that the VC-dimension of a hypothesis set (d_{VC}) is n . This means:
 - (a) There are n points we cannot shatter
 - (b) There are $n+1$ points we cannot shatter
 - (c) We cannot shatter any set of n points
 - (d) We cannot shatter any set of $n+1$ points
- (iv) How many dichotomies can you list on 4 points so that no subset of 2 is shattered?
 - (a) 4
 - (b) 5
 - (c) 6
 - (d) 3
- (v) For a neural network, which one of these structural assumptions is the one that most affects the trade-off between under-fitting and over-fitting?
 - (a) Number of hidden nodes
 - (b) Learning rate
 - (c) Initial choice of weights
 - (d) Use of constant term unit input
- (vi) Multi-layer perceptron can learn
 - (a) NAND
 - (b) XOR
 - (c) Both (a) and (b)
 - (d) Neither (a) nor (b)

- (vii) The back-propagation algorithm learns a globally optimal neural network with hidden layers
 (a) Always true (b) Always false
 (c) Mostly true (d) Mostly false
- (viii) After SVM learning, each Lagrange multiplier α_i takes either zero or non-zero value. What does it indicate in each situation?
 (a) A non-zero α_i indicates the data point i is a support vector, meaning it touches the margin boundary
 (b) A non-zero α_i indicates that the learning has not yet converged to a global minimum
 (c) A zero α_i indicates that the data point i has become a support vector data point, on the margin
 (d) A zero α_i indicates that the learning process has identified support for vector i
- (ix) Consider the following data set, where X_1 and X_2 are input and Y is the output class.
- | | | | | |
|-------|----|----|----|----|
| X_1 | 0 | 0 | 1 | 1 |
| X_2 | 0 | 1 | 0 | 1 |
| Y | +1 | -1 | -1 | +1 |
- Minimum training error can be achieved by
 (a) 3NN Classifier (b) SVM (Quadratic Kernel)
 (c) Both (a) and (b) (d) Neither (a) nor (b)
- (x) Which of the following is indicated by a hard margin in SVM?
 (a) The SVM allows very low error in classification of training data
 (b) The SVM allows high amount of error in classification of training data
 (c) The SVM allows no error in classification of training data
 (d) None of the above

Fill in the blanks with the correct word

- (xi) The growth function $h(N)$ for positive intervals ($h(X) = 1$ when $a \leq X \leq b$ and $h(X) = -1$ otherwise) is _____.
- (xii) In linear regression the _____ the R^2 value, the better the model fits your data.
- (xiii) _____ refers to a model that can neither model the training data nor generalize to new data.
- (xiv) The parameter C in SVM signifies _____.
- (xv) A perceptron can have _____ number of hidden layers.

Group – B

2. (a) Explain the contexts where linear regression is used. Write the linear regression algorithm, in detail. [[CO2](Apply/IOCQ)]
- (b) Illustrate a simple learning model using the concept of input, output, learning algorithm and hypothesis set. [[CO2](Remember/LOCQ)]
- (c) Briefly explain the difference between input space, feature space, and output space. [[CO2](Apply/IOCQ)]

(2 + 4) + 3 + 3 = 12

3. (a) Describe the Perceptron Learning Algorithm (PLA) and briefly explain the working principle of the algorithm. [[CO2](Apply/IOCQ)]
- (b) Classes attended by 10 students in machine learning and marks obtained in the examination are provided in the following table. Estimate the marks a student may obtain in the examination when she attended 20 classes, using linear regression.

Sl No	Attendance	Marks	Sl No	Attendance	Marks
1	28	43	6	28	39
2	27	39	7	26	36
3	23	27	8	21	36
4	27	36	9	22	31
5	24	34	10	28	37

[[CO2](Apply/IOCQ)]

$$(4 + 2) + 6 = 12$$

Group - C

4. (a) Calculate growth function and break point for (i) $h(x) = 1$ for $x \geq a$ and $h(x) = -1$ for $x < a$ and (ii) convex sets for N points. [[CO3](Apply/IOCQ)]
- (b) Define the mathematical definition of training and testing in context of Hoeffding's inequality. [[CO4](Apply/IOCQ)]
- (c) Decompose out-of-sample error in terms of bias and variance based on squared error measure. [[CO4](Apply/IOCQ)]
- $$4 + 4 + 4 = 12$$
5. (a) State and explain the Vapnik–Chervonenkis (VC) inequality in context of machine learning. [[CO4](Analyse/HOCQ)]
- (b) Define the VC dimension and discuss the significance of VC dimension in developing a machine learning model. [[CO3](Apply/IOCQ)]
- (c) Describe the relation between the number of observed data points and VC dimension of a model. [[CO3](Apply/IOCQ)]
- $$5 + 4 + 3 = 12$$

Group - D

6. (a) Assume we have a set of data from patients who have visited Heritage hospital during the year 2017. A set of features (e.g., temperature, height) have also been extracted for each patient. Our goal is to decide whether a new visiting patient has diabetes, heart disease, or Alzheimer (a patient can have one or more of these diseases). We have decided to use a neural network to solve this problem. We have two choices: (i) either to train a separate neural network for each of the diseases or (ii) to train a single neural network with one output neuron for each disease, but with a shared hidden layer. Which method do you prefer? Justify your answer. [[CO5](Analyse/HOCQ)]
- (b) Briefly explain the momentum and how is it being incorporated in the back propagation learning technique. [[CO5](Understand/LOCQ)]
- $$10 + 2 = 12$$

7. (a) Briefly explain the convolution, pooling and fully connected layers in a convolutional neural network. [[CO5](Understand/LOCQ)]
- (b) An input of volume $48 \times 48 \times 3$ is fed to a convolutional neural network. What would be the output volume of a convolution layer when you apply 8 (eight) $5 \times 5 \times 3$ filters with stride 2 and a zero-padding of size 1. Also calculate the number parameters involved due to this layer. [[CO5](Apply/IOCQ)]
- $6 + (4 + 2) = 12$**

Group - E

8. (a) Describe the concept of regularization and over-fitting in machine learning. [[CO1](Understand/LOCQ)]
- (b) A linearly separable dataset is given in the following Table. Predict the class of (0.6, 0.8) using a support vector machine classifier.

X_1	X_2	Y	Lagrange Multiplier
0.3	0.4	+1	5
0.7	0.6	-1	8
0.9	0.5	-1	0
0.7	0.9	-1	0
0.1	0.05	+1	0
0.4	0.3	+1	0
0.9	0.8	-1	0
0.2	0.01	+1	0

[[CO6](Apply/IOCQ)]
 $(4 + 4) + 4 = 12$

9. (a) A linearly separable dataset is given in the table below. Predict the class of (0.6, 0.8) using a support vector machine classifier. Show all the relevant computations.

X_1	X_2	Y	Lagrange Multiplier
0.3858	0.4687	+1	65.5261
0.9218	0.4103	-1	0
0.7382	0.8936	-1	0
0.1763	0.0579	+1	0
0.4057	0.3529	+1	0
0.9355	0.8132	-1	0
0.2146	0.0099	+1	0
0.4871	0.611	-1	65.5261

[[CO6](Analyse/HOCQ)]

12

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	11.46	60.42	28.12