MACHINE LEARNING
(INFO 4122)

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

Figures out of the right margin indicate full marks.

Candidates are required to answer Group A and
<u>any 4 (four)</u> from Group B to E, taking <u>one</u> 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) Logistic regression is used when you want to _____.
(a) predict a dichotomous variable from continuous or dichotomous variables
(b) predict a continuous variable from dichotomous variables
(c) predict any categorical variable from several other categorical variables
(d) predict a continuous variable from dichotomous or continuous variables

(ii) Back propagation is a learning technique that adjusts weights in the neural network by propagating weight changes _____.
(a) forward from source to sink
(b) backward from sink to source
(c) forward from source to hidden nodes
(d) backward from sink to hidden nodes

(iii) For a neural network, which one of these structural assumptions is the one that most affects the trade- off between under-fitting and overfitting?
(a) Number of hidden nodes   (b) learning rate
(c) Initial choice of weights   (d) Use of constant term unit input.

(iv) 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.

(v) Statement 1: The error surface followed by the gradient descent back propagation algorithm changes if we change the training data.
Statement 2: Stochastic gradient descent is always a better idea than batch gradient descent.
(a) only statement 1 is true   (b) only statement 2 is true
(c) both are true   (d) both are false.

(vi) 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 _____.
(a) overfitting   (b) generalizing
(c) regularizing   (d) none of the above.

(vii) The effectiveness of an SVM depends upon _____.
(a) selection of kernel   (b) kernel parameters
(c) soft margin parameter C   (d) all of the above.

(viii) Which of the following is/are true for RBF network?
I. The network training is divided into two stages: first the input vectors are non-linearly mapped and then linearly classifier is trained.
II. The training/learning is slow as compared to MLP.
(a) Only I   (b) Only II   (c) Both I and II   (d) None of these

(ix) Which of the following layer of CNN controls the overfitting problem?
(a) Convolutional Layer   (b) Hidden Layer
(c) Down Sampling Layer   (d) Fully Connected Layer.

(x) Perceptron can learn _____.
(a) AND   (b) XOR   (c) both (a) and (b)   (d) none of these

Fill in the blanks with the correct word

(xi) When zero-padding is not added, it is called _____.

(xii) Machine Learning is _____ of Artificial Intelligence.

(xiii) The generalization error in terms of support vector machine is _____.

(xiv) _____ layer of CNN controls the over fitting problem.

(xv) The minimum time complexity for training an SVM is _____ .

Group - B

2. (a) Define Precision and Recall in the context of classification.
   (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.
   (c) Distinguish between ID3 and C4.5 decision tree algorithm.

   4 + 4 + 4 = 12

3. (a) Discuss linear regression with example.
   (b) Discuss logistic regression with example.

   6 + 6 = 12

Group - C

4. (a) Describe Linear Discriminant Analysis (LDA) in the context of dimensionality reduction.
   (b) Explain the concept of non hierarchical clustering.
   (c) Describe Fuzzy C-Means clustering algorithms.

   4 + 3 + 5 = 12

5. (a) What is a kernel in SVM? Why do we use kernels in SVM?
   (b) Explain soft margin and hard margin Support Vector Machine.

   (3 + 3) + 6 = 12

Group - D

6. (a) What do you mean by cost function and loss function?
   (b) Why RelU (Rectified Linear Unit) is most popular activation function?
   (c) Briefly explain the momentum and how is it being incorporated in the back propagation learning technique.

   (2 + 2) + 4 + 4 = 12

7. (a) Explain different types of Gradient Descent in detail.
   (b) Explain forward and backward propagation in neural network model in detail.

   6 + 6 = 12

Group - E

8. (a) Justify how a CNN can identify an object even when it is transformed.
   (b) Explain the architecture of the AlexNet deep convolutional network.

   6 + 6 = 12

9. (a) Discuss max-pooling and sum-pooling in the context of convolution.
   (b) Discuss SoftMax function in the context of Deep Learning in detail.

   6 + 6 = 12