MACHINE LEARNING
(CSEN 3233)

Time Allotted : 3 hrs **Full Marks : 70**

Figures out of the right margin indicate full marks.

Candidates are required to answer Group A and any 5 (five) from Group B to E, taking at least one from each group.

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

Group – A
(Multiple Choice Type Questions)

1. Choose the correct alternative for the following: **10 × 1 = 10**

(i) Which of the following hypothesis cannot be classified using a single perceptron in 2 dimensional input space?
(a) Positive interval (b) AND (c) Positive ray (d) OR

(ii) Historical data which is used for learning in a supervised learning model is called
(a) Training data set (b) Test data set
(c) Validation data set (d) none of the above.

(iii) In a neural network with (L +1) numbers of layers which includes the input (0^{th} layer), output (L^{th} layer) and hidden layers, the term "Feed Forward" refers to the processing of input signal at each perception of the network to be performed layer-wise in the order
(a) layer i=L-1 to i=0 (b) layer i=0 to i=L-1
(c) layer I =1 to i=L (d) layer i=L to i=0.

(iv) 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.

(v) The maximum number of dichotomies which can be achieved by of a positive ray hypothesis using a training data set having 10 data points is
(a) 10 (b) 11 (c) 9 (d) 12.

(vi) 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.

(vii) In an MLP, the number of nodes in the input layer is 5 and the hidden layer is 3 excluding the bias node in the hidden layer). The maximum number of connections from the input layer to the hidden layer are
(a) Less than 15 (b) 15
(c) More than 15 (d) It is an arbitrary value.

(viii) Which of the following functions can be used in logistic regression model?
(a) cosh (b) ReLu (c) Sigmoid (d) Tanh.

(ix) The objective of a Linear Support Vector Machine is to classify linearly separable data points:
(a) using the hyper plane with maximum margin
(b) using the hyper plane with minimum margin
(c) using any hyper plane which can separate the data points
(d) All of the above.

(x) 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.

Group – B

2. (a) What is linear regression? Describe the working principle of linear regression for N input training data points/vectors, where the dimension of each input vector i.e., $d \geq 1$.

(b) Briefly discuss about the various components of learning for non-deterministic target with the help of a block diagram.

(1 + 5) + 6 = 12

3. (a) Explain the conditions under which the following learning rate should be used in gradient descent technique:
(i) extremely small learning rate value
(ii) moderately large learning rate value.

(b) Explain how stochastic gradient descent will have better performance compared to batch gradient descent technique.

(c) What is the error function used for misclassification error in single perceptron learning model? Derive the weight update equation of single perceptron model.

3 + 4 + 5 = 12

Group – C

4. (a) Derive the growth function of n input data points for the following hypotheses:
(i) Positive intervals
$f(x) = +1$ for $a \leq x \leq b$;

Otherwise, f(x) = -1

(ii) Positive ray f(x) = +1 for x>=a

Otherwise, f(x) = -1

(b) Suppose 4 input data points X_1, X_2 ,X_3, and X_4 need to be classified into two clases +1 and -1. Determine a set or list of maximum possible dichotomies for these 4 data points, which can be obtained under the following three conditions:
(i) break point is 2 (ii) break point is 3. (iii) break point is 4.

(3 + 3) + 6 = 12

5. (a) Explain the concept of dichotomy, growth function, break point andf VC dimension in machine learning with the help of a suitable example.

(b) Explain the derivation of the recurrence relation that defines B(N,K) , which denotes the maximum number of dichotomies for N data points with k break points.

6 + 6 = 12

Group – D

6. (a) Explain the derivation of weight update equation of a feed forward Multi Layered Perceptron model (i.e., Neural network), based on back propagation learning algorithm.

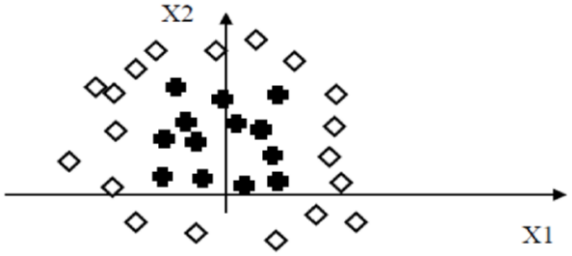
(b) Describe all the matrices (or 2D arrays) along with their respective dimensions, which are required for implementation(/coding) of a feed forward neural network based on back propagation learning algorithm, using a training data set of N input vectors having dimension d >=1.

7 + 5 = 12

7. (a) Briefly discuss about the cross entropy error function and its usage in logistic regression model.

(b) Briefly discuss about the differences between soft threshold and hard threshold activation function with suitable examples.

(c) Suppose that we want to build a neural network that classifies two dimensional data (i.e., X = [x_1, x_2]) into two classes: white diamonds and black circles. We have a set of training data that is plotted as follows:



Draw a neural network that can solve the above classification problem. Justify your choice of the number of nodes and the architecture.

4 + 3 + 5 = 12

Group – E

8. (a) Explain the meaning of the term "support vector" in SVM ?

   (b) Construct the primal problem and then derive the dual formulation using Lagrangian multiplier for the optimization problem, as defined by linear SVM classifier.

   (c) Discuss how an unknown out of sample data point will be classified by a hard margin linear SVM classifier using a set of non zero Langrange multiplier/solution, which were returned by a quadratic programming solver.

2 + 6 + 4 = 12

9. (a) Briefly explain the Radial Basis function and its application in RBF Network

   (b) Briefly explain how SVM classify non linearly separable data points by using kernel function.

6 + 6 = 12

Department & Section	Submission Link
CSE	https://classroom.google.com/c/MzAwODI2OTk3Njk3/a/MzY0NTM5ODc3Mzc5/details