MACHINE LEARNING
(CSEN 5131)

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) Consider the following data set, where A and B are input.

A	0	0	1	1
B	0	1	0	1
output	-1	+1	-1	-1

Minimum training error can be achieved by

(a) Single Layer Perceptron (b) SVM (Quadratic Kernel)

(c) Both (a) and (b) (d) Neither (a) nor (b)

(ii) We have seen two error functions that are used for neural networks: sum-of-squares error (SSE), and cross-entropy error (CEE). Suppose we are training a neural network for binary classification. Which of the following are true?

(a) SSE cannot be used; it works only for regression.

(b) CEE should be preferred to SSE, because CEE is closer to classification error, which is what we really care about.

(c) CEE should be preferred to SSE, because CEE also takes into account the magnitude of error, rather than just right/wrong.

(d) Both CEE and SSE can give good results, but in principle CEE might be slightly preferable because it corresponds to maximizing the likelihood of the data.

(iii) Multi Layered Perceptron model is used to classify

(a) Linearly separable classes of data

(b) Non-linearly separable classes of data

(c) Both (a) and (b)

(d) None of these.

(iv) Growth function of a hypothesis H is n + 1, means the VC dimension $d_{VC}(H)$ satisfies

(a) $d_{VC}(H) = n + 1$ (b) $d_{VC}(H) > n + 1$

(c) $d_{VC}(H) = n$ (d) No conclusion.

(v) *H* consists of all hypotheses in two dimensions *h*: $R^2 \rightarrow \{-1, +1\}$ that are positive inside some convex set and negative elsewhere. The growth function h(N) of *H* is
(a) N (b) N+1 (c) ∞ (infinity) (d) 2^N

(vi) This technique uses mean and standard deviation scores to transform real-valued attributes.
(a) decimal scaling (b) min-max normalization
(c) z-score normalization (d) logarithmic normalization.

(vii) Suppose your model is overfitting. Which of the following is NOT a valid way to try and reduce the overfitting?
(a) Increase the amount of training data
(b) Improve the optimization algorithm being used for error minimization
(c) Decrease the model complexity
(d) Reduce the noise in the training data.

(viii) Using the hypothesis H, we can shatter n points. This implies that
(a) $d_{VC} = n$ (b) $d_{VC} \geq n$
(c) $d_{VC} \leq n$ (d) No conclusion
d_{VC} being the VC-dimension of H.

(ix) Regarding bias and variance, which of the following statements are true? (Here `high' and `low' are relative to the ideal model.)
(a) Models which overfit have a high bias
(b) Models which overfit have a low bias
(c) Models which underfit have a high variance
(d) Models which underfit have a low variance.

(x) Data used to optimize the parameter settings of a supervised learning model
(a) training (b) testing
(c) verification (d) none of the above.

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) Marks obtained by 10 students in the class test and semester examination in machine learning are provided in the following table. Estimate the marks a student may obtain in the semester examination when she got 20 in class test, using linear regression.

Sl No	Class Test Marks	Semester Marks	Sl No	Class Test Marks	Semester Marks
1	28	63	6	28	51

2	27	49	7	26	66
3	23	43	8	21	36
4	17	36	9	22	31
5	24	39	10	19	37

2 + 2 + 8 = 12

3. (a) Discuss the back propagation learning algorithm for a single layer artificial neural network using an appropriate example.

   (b) How stochastic gradient descent differs from standard gradient descent?

   (c) "Gradient descent technique always ensures global optima"— Argue in favour or against this statement.

8 + 2 + 2 = 12

Group – C

4. (a) Define Dichotomy, Growth function and Break point

   (b) Calculate growth function and break point for (i) positive rays, (ii) positive intervals and (iii) convex set for N points.

   (c) Proof that VC dimension for perceptron is d+1, where d is the dimension of the data points under consideration.

3 + 3 + 6 = 12

5. (a) Find the growth function for positive ray based classification.

   (b) You are given 3 points X_1, X_2 and X_3. Calculate the number of dichotomies for the following two cases:
   (i) break point is 3 (ii) break point is 2.

2 + (5 + 5) = 12

Group – D

6. Construct the primal problem and then derive the Lagrangian and its Dual for the optimization problem as defined by linear SVM – classification.

12

7. (a) Define Information gain and gain in the Gini index.

   (b) Consider the following data set for a binary class problem.

Sl No	A	B	Gender
1	T	F	C1
2	T	T	C1
3	T	T	C1
4	T	F	C2
5	T	T	C1
6	F	F	C2

7	F	F	C2
8	F	F	C2
9	T	T	C2
10	T	F	C2

(i) Calculate the information gain when splitting on A and B. Which attribute would the decision tree induction algorithm choose?

(ii) Calculate the gain in the Gini index when splitting on A and B. Which attribute would the decision tree induction algorithm choose?

2 + (5 + 5) = 12

Group – E

8. (a) Describe the k-means algorithm.

(b) Perform K-means clustering on all the points in the following table, where K=2. Randomly select the initial seeds and perform the algorithm for two iterations.

Points	X co-ordinate	Y co-ordinate
p1	1	9
p2	2	10
p3	7	4
p4	10	3
p5	5	6
p6	6	11
p7	3	4
p8	4	9
p9	8	1
p10	3	12
p11	7	6
p12	11	2

3 + 9 = 12

9. (a) Briefly discuss the significance of pre-defined parameter required in db-scan clustering algorithm.

(b) Apply DBSCAN on the following set of data points, to determine the clusters. Show all steps in detail, assuming *eps* = 1 and *minPts* = 2.
Data Points: (3,-1), (3,4), (2,3), (4,3), (4,-2), (10,1), (10,2), (20,11), (20,12), (21,11), (21,12).

3 + 9 = 12

Department & Section	Submission Link
CSE	https://classroom.google.com/c/MjIxNzQ1Njc2MzQ3/a/MjkxNDEwNTE1Mjc3/details