MACHINE LEARNING
(CSEN 4264)

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 a binary classification problem. Suppose I have trained a model on a linearly separable training set, and now I get a new labeled data point which is correctly classified by the model, and far away from the decision boundary. If I now add this new point to my earlier training set and re-train, in which cases is the learnt decision boundary likely to change?

(a) When my model is a perceptron (b) When my model is logistic regression
(c) When my model is an SVM (d) None of the above.

(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 Layer Perceptron can be used as

(a) a classifier (b) an estimator
(c) none of (a) and (b) (d) both (a) and (b).

(iv) Growth function of a hypothesis H is n + 1, means the VC dimension dVC(H) satisfies:

(a) dVC(H) = n + 1 (b) dVC(H) > n + 1
(c) dVC(H) = n (d) No conclusion.

(v) ***H*** consists of all hypotheses in two dimensions ***h*: *R*2 → {-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) What is back propagation in ANN?
(a) It is another name given to the curvy function in the perceptron.
(b) It is the transmission of error back through the network to adjust the inputs
(c) It is the transmission of error back through the network to allow weights to be adjusted so that the network can learn
(d) It is a feed forward technique.

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) Derive the linear regression formula for single dependent variables.

(d) 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 + 4 + 4 = 12

3. (a) Discuss with example the in-sample error and out-of-sample error.

   (b) Write the error functions used in Perceptron Learning Algorithm (PLA), Regression and Logistic Regression.

   (c) Briefly explain the contexts where logistic regression is used. Write the logistic regression algorithm, in detail.

2 + 3 + (2 + 5) = 12

Group – C

4. (a) Explain the importance of VC dimension in machine learning.

   (b) Find the VC Dimension for the following hypotheses:
   Positive intervals
   - $F(x) = +1$ for $a \leq x \leq b$; -1 otherwise.
   - Perceptron in R2.

   (c) Show that the growth function of a hypothesis is polynomial when there exists a finite break point.

3 + 3 + 6 = 12

5. (a) Explain the Bias-Variance trade off in the context of learning.

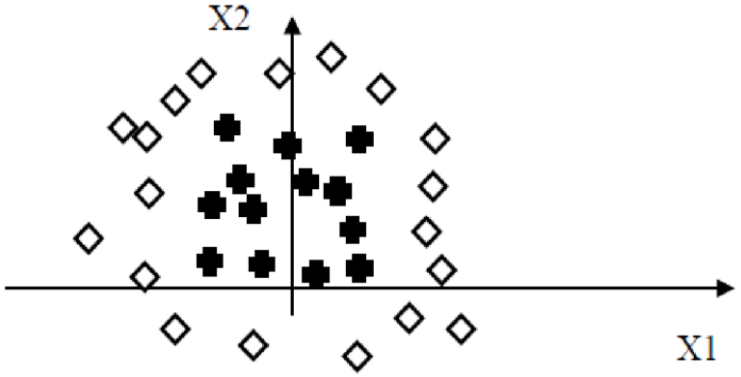
   (b) Write short notes on any two of the followings:
   (i) Overfitting, (ii) Regularization and (iii) Validation.

6 + (2 × 3) = 12

Group – D

6. (a) Explain the multi-layer perceptron (ANN with one hidden layer) for classification using the back propagation algorithm for a data set D consisting of the training tuples and their associated target values.

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



Draw a network that can solve this classification problem. Justify your choice of the number of nodes and the architecture. Draw the decision boundary that your network can find on the diagram.

8 + 4 = 12

7. (a) Briefly explain the convolution, pooling and fully connected layers in a convolutional neural network.

   (b) An input of volume 48 × 48 × 3 is fed to a Convolutional Neural Network. What would be the output volume of a convolution layer when you apply 8 (eight) 5 × 5 × 3 filters with stride 2 and a zero(0) padding of size 1. Also calculate the number parameters involved due to this layer.

6 + (4 + 2) = 12

Group – E

8. (a) Construct the primal problem and then derive the Lagrangian for the optimization problem as defined by linear SVM – classification.

   (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. Show all the relevant computations.

X_1	X_2	Y	Lagrange Multiplier
0.3858	0.4687	+1	65.5261
0.4871	0.611	-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

8 + 4 = 12

9. (a) Explain how kernel function is used in non-linear support vector machines. Also justify the statement that "One can use infinite-dimensional spaces with the kernel trick" in the perspective of non-linear SVM classification.

   (b) When a polynomial kernel of degree 2 ($K(x,y) = (xy+1)^2$) is used and C is set to 100, we get the Lagrange multipliers as follows:
   $\alpha_1=0$, $\alpha_2=2.5$, $\alpha_3=0$, $\alpha_4=7.333$, $\alpha_5=4.833$

Identify the support vectors and derive the discrimination function.

6 + 6 = 12

Department & Section	Submission Link
CSE A + B + C	https://classroom.google.com/c/MzAwMzI0MjE0OTYw/a/MzU4OTQ3MzU5MjYz/details