B.TECH/CSE/8TH SEM/CSEN 4264/2018

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) An artificial neuron receives n inputs $x_1, x_2, \ldots\ldots, x_n$ with weights $w_1, w_2, \ldots\ldots, w_n$ attached to the input links. The weighted sum ____________ is computed to be passed on to a non-linear filter Φ called activation function to release the output.
(a) $\Sigma\, w_i$ (b) $\Sigma\, x_i$ (c) $\Sigma\, w_i + \Sigma\, x_i$ (d) $\Sigma\, w_i{*}\, x_i$.

(ii) The growth function h(N) for positive rays (h(X) = 1 when X > a and h(X) = -1 otherwise) is
(a) N+1 (b) N (c) 2^N (d) ∞ (Infinity).

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

(iv) 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.

(v) Which 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

(vi) 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.

(vii) 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.

(viii) **Statement- 1:** The SVM learning algorithm is guaranteed to find the globally optimum hypothesis with respect to its object function.
Statement- 2: The VC dimension of a perception is smaller than VC dimension of a linear SVM.
(a) Only statement- 1 is true (b) Only statement -2 is true
(c) Both are true. (d) Both are false.

(ix) **Statement-1:** Kernel density estimator can be used to perform classification.
Statement -2: Cross validation can be used to select the number of iterations in boosting; this procedure may help reduce over fitting.
(a) Only Statement -1 is true (b) Only statement- 2 is true
(c) Both are true (d) Both are false.

(x) Consider the following data set, where X1 and X2 are input and Y is the output class.

X1	0	0	1	1
X2	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).

Group – B

2. (a) Write short notes on any two of the followings:
   (i) Hoeffding's inequality in learning theory
   (ii) Error vs. Noise
   (iii) Bias-Variance tradeoff

   (b) Write the Perceptron Learning Algorithm (PLA) and briefly explain the working principle of the algorithm.

6 + 6 = 12

3. (a) Derive the linear regression formula for multiple dependent variables.

   (b) Explain briefly the difference between input space and feature space. Also explain how the derived linear regression formula can be used for nonlinear cases.

6 + 6 = 12

Group – C

4. (a) Explain the importance of VC dimension in machine learning.
   (b) "Sum of square error is not always suitable." Justify the statement with appropriate example. Explain the error and noise in the context of learning.
   (c) Explain the Bias-Variance trade- off in the context of learning.

4 +4 + 4 =12

5. (a) Define growth function and break point.
   (b) Find the growth function and break point for the positive intervals
   F(x) = +1 (when a ≤ x ≤ b); otherwise, -1.
   (c) Consider a hypothesis class H as a set of all perceptrons in 2D plane, i.e., $H=\{h_k \mid h_k = f(w_0+w_1x_1+w_2x_2)$, where f(x)=1, iff x>=0, otherwise f(x)=0}
   Find the VC dimension of this hypothesis H.

3 + 3 + 6 = 12

Group – D

6. (a) Derive the weight update equation of a feed forward multi-layered Perceptron network using back propagation learning.

   (b) Briefly explain the momentum and how is it being incorporated in the back propagation learning technique.

10 + 2 = 12

7. (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).

B.TECH/CSE/8TH SEM/CSEN 4264/2018

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.

   (b) Explain how MLP can be used as an estimator.

8 + 4 = 12

Group – E

8. 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
0.4871	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

12

9. Sometimes data is just nonlinearly separable or data has errors and one wants to ignore them to obtain a better solution. In fact, this is achieved by relaxing the margin, in other words, using a soft margin.

   Derive the Lagrangian for the optimization problem as defined by linear SVM – soft margin classification.

12