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
<u>any 5 (five)</u> from Group B to E, taking <u>at least one</u> 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 $x_1, x_2, ..., x_n$ with weights $w_1, w_2, ..., w_n$ 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) Σw_i (b) Σx_i (c) $\Sigma w_i + \Sigma x_i$ (d) $\Sigma w_i * x_i$

    (ii)   Which of the following methods do we use to best fit the data in Logistic Regression?
            (a) Least Square Error   (b) Maximum Likelihood
            (c) Jaccard distance   (d) Both (a) and (b).

    (iii)   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.

    (iv)   Which of the following statement(s) is / are true for Gradient Decent (GD) and Stochastic Gradient Decent (SGD)?
            1. In GD and SGD, you update a set of parameters in an iterative manner to minimize the error function.
            2. In SGD, you have to run through all the samples in your training set for a single update of a parameter in each iteration.
            (a) Only 1   (b) Only 2   (c) Both 1 and 2   (d) None of 1 and 2.

    (v)   What is generalization?
            (a) The ability of a pattern recognition system to approximate the desired output values for pattern vectors which are not in the test set.
            (b) The ability of a pattern recognition system to approximate the desired output values for pattern vectors which are not in the training set.
            (c) Can be either way
            (d) None of the mentioned.

(vi)   Perceptron cannot learn
        (a) AND   (b) XOR   (c) both (a) and (b)   (d) none of these.

(vii)   "Convolutional networks have generally more parameters than their equivalent
        fully connected networks"- This Statement is _____.
        (a) always true   (b) always false   (c) mostly true   (d) mostly false

(viii) 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.

(ix)   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.

(x)   Which of the following gives non-linearity to a neural network.
        (a) Convolution operator (weighted sum of the inputs)
        (b) Stochastic gradient descent
        (c) Sigmoid activation
        (d) Non-zero bias.

Group- B

2. (a)   Discuss with example the in-sample error and out-of-sample error.
                                          [(CO2)(Remember/LOCQ)]
    (b)   Briefly explain error and noise with example.   [(CO3)(Understand/LOCQ)]
    (c)   Classes attended by 10 students
          in machine learning and marks
          obtained in the examination are
          provided in the Table 1.
          Estimate the marks a student
          may obtain in the examination
          when she attended 20 classes,
          using linear regression.
          [(CO4)(Analyse/IOCQ)]

Table 1

Sl No	Attendance	Marks
1	28	43
2	27	39
3	23	27
4	27	36
5	24	34
6	28	39
7	26	36
8	21	36
9	22	31
10	28	37

4 + 6 + 2 = 12

3. (a)   Derive the linear regression formula for single dependent variables.
                                          [(CO4)(Remember/LOCQ)]
    (b)   What is error in prediction? Explain, with appropriate examples, three different
          error measures used to deal with different prediction problems.
                                          [(CO5)(Describe/HOCQ)]

(c)   Write the Perceptron Learning Algorithm (PLA) and briefly explain the working principle of the algorithm.   [(CO2)(Analyze/IOCQ)]

4 + 4 + 4 = 12

Group - C

4.   (a)   Explain the importance of VC dimension in machine learning?
                                                                      [(CO4)(Remember/HOCQ)]
     (b)   Find the VC dimension for the following hypotheses:
             (i)   Positive intervals F(x) = +1 for a ≤ x ≤ b; -1 otherwise.
             (ii) Perceptron in R2.   [(CO2)(Analyze/LOCQ)]
     (c)   Explain the Bias-Variance trade off in the context of learning.
                                                                      [(CO3)(Analyze/IOCQ)]

3 + 3 + 6 = 12

5.   (a)   Define Dichotomy, Growth function and break point.
                                                                      [(CO3)(Understand/LOCQ)]
     (b)   Calculate growth function and break point for positive intervals for N points.
                                                                      [(CO4)(Analyze/IOCQ)]
     (c)   You are given 3 points X_1, X_2 and X_3. Calculate the number of dichotomies when (i) break point is 3 and (ii) break point is 2.   [(CO6)(Apply/HOCQ)]

3 + 3 + 6 = 12

Group - D

6.   (a)   What does the learning rate do in Back-Propagation training?
                                                                      [(CO4)(Remember/LOCQ)]
     (b)   Describe what is likely to happen when a learning rate is used that is too large, and when one is used that is too small. How can one optimize the learning rate?
                                                                      [(CO3)(Understand/LOCQ)]
     (c)   Describe the importance of using bias in neural network.
                                                                      [(CO2)(Remember/LOCQ)]
     (d)   Explain the main reasons why a Back-Propagation training algorithm might not find a set of weights which minimizes the training error for a given feed-forward neural network.   [(CO1)(Analyze/IOCQ)]
     (e)   Explain the purpose of the momentum term that is often included in the Back-Propagation learning algorithm.   [(CO1)(Analyze/IOCQ)]

2 + 3 + 2 + 3 + 2 = 12

7.   (a)   An input of volume 48×48×3 is fed to a Convolutional Neural Network. What would be the output volume when you apply?
             (i)   Four 5×5×3 filters with stride 1 without padding.
             (ii) Three 5×5×3 filters with stride 2 and a zero (0) padding of size 1.
                                                                      [(CO3)(Understand/IOCQ)]
     (b)   What is the difference between Recurrent Neural Network and Convolutional Neural Network?   [(CO1)(Remember/LOCQ)]

(c)   You are asked to simulate the Boolean function x1 ∧ x2 ∨ (¬$X3$) using a multi-layer perceptron. Construct the network and explain how your network is able to model the said function.   [(CO1)(Analyze/IOCQ)]

(2 + 2) + 2 + 6 = 12

Group - E

8.
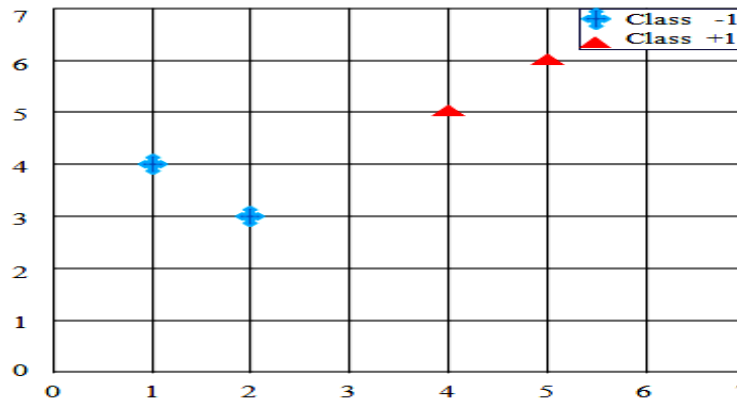
Fig.1

(a)   Support vector machines learn a decision boundary leading to the largest margin from both classes. You are training SVM on a tiny dataset with 4 points shown in Fig.1. This dataset consists of two examples with class label -1 (denoted with plus), and two examples with class label +1 (denoted with triangles).
(i) Find the weight vector w and bias b. What's the equation corresponding to the decision boundary?
(ii) Circle the support vectors and draw the decision boundary.

[(CO4)(Analyse/HOCQ)]

(b)   What is overfitting? How to deal with this situation?

[(CO4)(Remember/IOCQ)]

(c)   What is weight decay?   [(CO4)(Remember/LOCQ)]

(3 + 1) + (2 + 4) + 2 = 12

9.   Write short notes on any three of the followings: **(4 × 3) = 12**
(i)   Logistic Regression
(ii)   Kernel trick of SVM
(iii) Regularization in Machine Learning
(iv) Validation in Machine Learning.   [(CO1)(Remember/IOCQ)]

Cognition Level	LOCQ	IOCQ	HOCQ
Percentage distribution	32.29	19.79	47.92

Course Outcome (CO):

After the completion of the course students will be able to
1. Learn and understand the basics of machine learning approaches and paradigm.
2. Understand and describe various machine learning algorithms.
3. Understand complexity of Machine Learning algorithms and their limitations.

4. Mathematically Analyse various machine learning approaches and paradigms
5. Analyse various machine learning techniques to get an insight of when to apply a particular machine learning approach.
6. Apply common Machine Learning algorithms in practice and implementing their own using real-world data.

*LOCQ: Lower Order Cognitive Question; IOCQ: Intermediate Order Cognitive Question; HOCQ: Higher Order Cognitive Question.