B.E. COMPUTER SCIENCE AND ENGINEERING FOURTH YEAR FIRST SEMESTER - 2018

Machine Learning

Time: Three hours

Full Marks: 100

*Answer any **FIVE** questions.*

1. **a)** How a hypothesis is represented for linear unit perceptron model? Derive gradient descent rule for a sigmoid unit perceptron model?

    b) What is the drawback of gradient descent search procedure? What is the role of learning rate and feature scaling while training artificial neural networks?   10 + (5 +5) = **20 marks**

2. a) Differentiate between lazy learning and eager learning. Discuss one lazy learning method with a suitable example.
    b) Write the Bayes theorem. Explain posterior probability and prior probability with examples. What are the problems in designing an unbiased Bayesian learning algorithm?

    (3+7)+ (2+5+3) = **20 marks**

3. The following is an Artificial Neural Networks, with sigmoid units in the hidden layer and the output layers. The weights have been set arbitrarily between all the units.
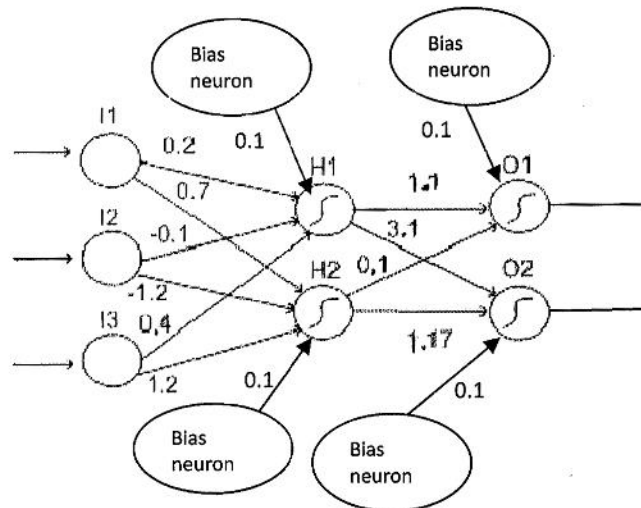
Consider that the following training example is submitted to the net (shown in the above figure).

X1	X2	X3	t1	t2
0.2	0.5	0.3	0.9	0.1

In the above training example, <x1, x2, x3> is the input vector and <t1, t2> is the target vector
Now show the forward pass to compute responses at nodes H1, H2 and O1, O2. Then use backpropagation algorithm to find updates for weights associated with the connections between hidden and output layers. Assume the learning rate η = 0.2. You may assume the values of the other parameters if necessary.   **20 marks**

[Turn over

4. **a)** *NASA* wants to be able to discriminate between the species: Martians (M) and Humans (H) based on the following characteristics: Green∈{N, Y}, Legs∈{2, 3}, Height∈{S, T}, Smelly∈{N, Y}. Our available training set (S) is as follows:

	Species	Green	Legs	Height	Smelly
1)	M	N	3	S	Y
2)	M	Y	2	T	N
3)	M	Y	3	T	N
4)	M	N	2	S	Y
5)	M	Y	3	T	N
6)	H	N	2	T	Y
7)	H	N	2	S	N
8)	H	N	2	T	N
9)	H	Y	2		N
10)	H	N	2	T	Y

Compute information gain with respect to S and each of the attributes and find the best attribute among the attributes..

b) What is called overfitting? Why it occurs? Explain one method for avoiding overfitting in decision tree learning.   10 + (5 + 5) = **20 marks**

5. **a)** What are basic differences between Artificial Neural Networks and Support Vector Machines (SVM)?

    b) What is called support vectors? Show primal formulation (constrained optimization problem) for linear SVM. Explain with a suitable example how kernel methods are useful while applying SVM to non-linear data.   5 + (2+8+5) = **20 marks**

6. Write short notes on the following (**Any Two**)
    a) Naïve Bayesian Classifier
    b) Performance Measures for Machine Learning algorithms
    c) Effect of Momentum parameter in Artificial Neural Networks
    d) A brief introduction to WEKA   (10 x 2) = **20 marks**

7. What is unsupervised learning? What is clustering? Mention the convergence criteria of K-means clustering algorithm. Write the *K*-means clustering algorithm with a simple example. What are drawbacks of K-means clustering algorithm?   (3 +3 +3+8+ 3)= **20 marks**