Ref. No.: Ex/PG/MTCT/T/222B/2018

M.TECH. COMPUTER TECHNOLOGY SECOND YEAR SECOND SEMESTER EXAM 2018

MACHINE LEARNING

Time: Three hours

Full Marks: 100

Answer any FIVE questions.

1. Describe the Naïve Bayesian learning algorithm. Give a suitable example to show how it works. How continuous attributes are handled in Bayesian learning?

   What is overfitting? Give an example

   5 + 5 + 5 +5 = 20 marks

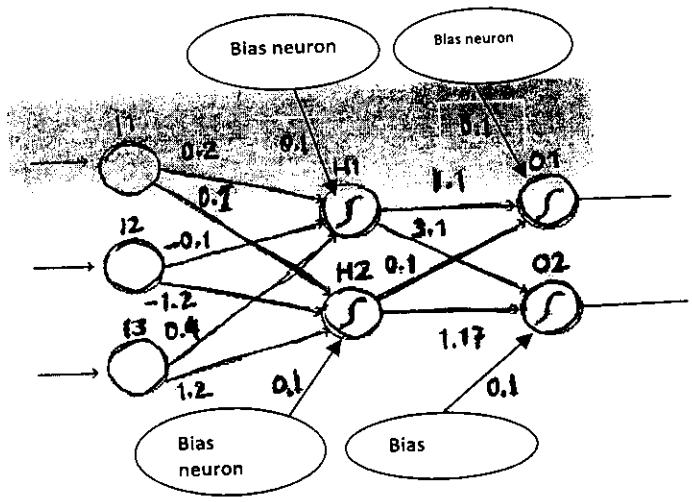
2. Consider the following set of training examples and show the decision tree that would be learned by ID3 algorithm. Compute the value of information gain for each candidate attribute at each step in growing the tree. Show how the obtained decision tree will classify the test instance *<sunny, mild, normal, weak>*.

Day	Outlook	Temp.	Humidity	Wind	Play Tennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Weak	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cold	Normal	Weak	No
D10	Rain	Mild	Normal	Strong	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

20 marks

3. 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

4. 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 $\eta = 0.2$. You may assume the values of the other parameters if necessary. 20 marks

5. a) How to measure purity (impurity) of a collection of training examples? Define information gain measure? What is its use in decision tree learning? Give a suitable example.

b) What is difference between shallow and deep neural networks? What is difference between Sigmoid and ReLu activation function? What are the advantages of ReLu activation over sigmoid activation function? When *deep neural network* is used for a classification problem, sigmoid activation is used for output nodes whereas ReLu or tanh activations are used in the hidden nodes. Explain why? (2+3) + 5 +(2+5+3) = 20 marks

6. a) Describe Linear SVM. Why kernelization of inputs is done for classification using SVM?
   b) Consider the following confusion matrix depicting the performance of a classification model and compute (i) accuracy, (ii) error rate, (iii) class wise F-measure, (iv) overall F-measure

Actual class\Predicted class	yes	no
yes	6954	46
no	412	2588

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

7. a) Describe K-means clustering algorithm with an example. Write the convergence criteria for K-means clustering algorithm. Give an example where the K-means clustering algorithm does not perform well.
   b) What is the significance of momentum parameter used in training artificial neural networks?

10 + (3+2) +5 = 20 marks

8. Write short notes on
   a. Handling continuous attributes in decision tree learning
   b. Methods for avoiding overfitting in decision tree learning
   c. Unbiased Baysian learning

7 + 7+ 6 =20 marks