

MASTER OF COMPUTER SCIENCE AND ENGINEERING EXAMINATION, 2017
(1ST SEMESTER)

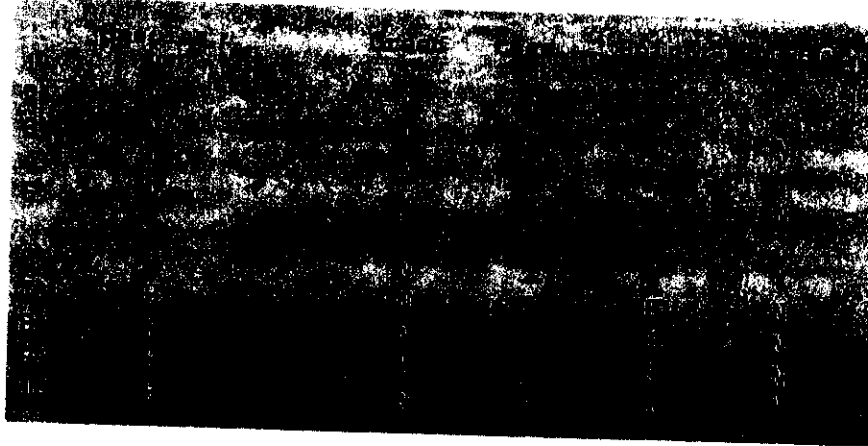
Machine Learning

Time: Three hours

Full Marks: 100

Answer any FIVE questions.

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



Develop a Naïve Bayes Classifier with the training data given above and predict species for the following test instances:

$\langle N, 3, T, Y \rangle$ and $\langle Y, 2, S, Y \rangle$

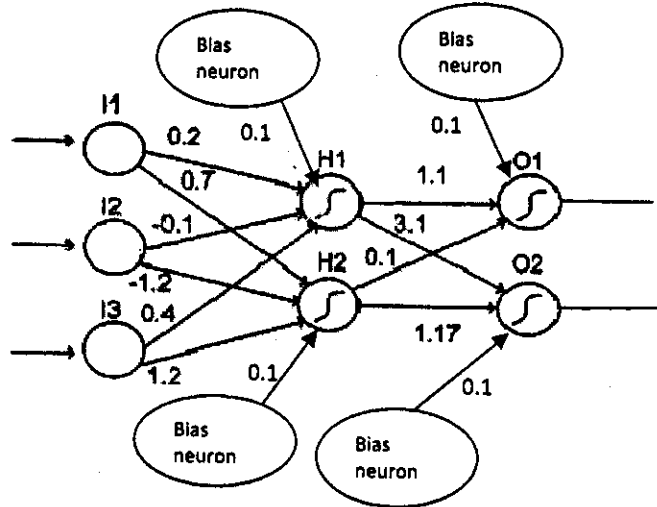
If the discrete attribute "Height" in the above table is changed to a continuous attribute and the column "Height" in the table is changed to $\langle 50, 80, 77, 55, 68, 70, 62, 90, 65, 85 \rangle$, then what would be prediction for the test instances, $\langle N, 3, 95, Y \rangle$ and $\langle Y, 2, 55, Y \rangle$. Show the calculation.

- b) If an unbiased Bayesian classifier is developed based on n attributes, how many parameters should be computed?
15 + 5 = 20 marks
2. a) How a hypothesis is represented by linear regression model when multiple features are used? Explain briefly.
- b) Define the cost function for linear regression with multiple features and compute the gradient of the cost function with respect to the j -th parameter, say θ_j .
- c) State *gradient descent algorithm* for learning parameters for linear regression model. What is the drawback of gradient descent search procedure?

5 + 10 + 5 = 20 marks

3. a) Explain the limitations of linear regression model which are overcome by logistic regression model.
 b) Write the cost function of logistic regression model. Explain how Logistic regression model can be applied for multi-class problem.
 c) What are the limitations of logistic regression model? 5+10+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	O1	O2
0.2	0.5	0.3	0.9	0.1

Now show the forward pass to compute responses at 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.

8 + 12 = 20 marks

5. a) What is overfitting? What causes overfitting? Explain one method which is used to avoid overfitting in decision tree learning?
 b) Explain with examples how continuous attributes and missing attributes are handled in decision tree learning. 10+10 =20 marks

6. a) What are basic differences between Artificial Neural Networks and Support Vector Machines (SVM)?
 b) What is the speciality about the Lagrange multipliers associated with the support vectors? Show primal formulation (constrained optimization problem) for linear SVM and derive also the dual formulation. Explain how kernel methods are useful while applying SVM to non-linear data

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

7. Consider the following set of training examples and show the decision tree that would be learned by ID3 algorithm. Clearly show 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

8. Write short notes on the following

- a) Polynomial Regression Model
- b) Performance Measures for Machine Learning algorithms
- c) K- fold cross validation
- d) Effect of Momentum parameter in Artificial Neural Networks

(5 x 4) = 20 marks

9. a) Describe K-means clustering algorithm with an example.

b) What is need of hidden layers in Artificial Neural Networks?

c) What are called controllable and uncontrollable parameters for machine learning algorithms?

Give example with respect to Artificial Neural Networks.

10 + 5+ 5= 20 marks