END SEMESTER EXAMINATION, DECEMBER, 2018

COURSE: M.Tech (1st Semester)
Course No: PGMTH1E011T
Course Title: Machine Learning

Time Allowed: Three Hours   **Max Marks:** 100

SECTION-A

Section A contains ten multiple choice questions and all are compulsory carrying 1.5 marks each

1. Choose the right options for each of the following multiple choice questions

I. Supervised learning and unsupervised clustering both require at least one

  A. hidden attribute.   B. output attribute.

  C. input attribute.   D. categorical attribute

II. Which statement is true about neural network and linear regression models?

  A. Both models require input attributes to be numeric.

  B. Both models require numeric attributes to range between 0 and 1.

  C. The output of both models is a categorical attribute value.

  D. Both techniques build models whose output is determined by a linear sum of weighted input attribute values.

III. Logistic regression is a _____ regression technique that is used to model data having a _____ outcome.

  A. linear, numeric   B. linear, binary

  C. nonlinear, numeric   D. nonlinear, binary

IV. The average squared difference between classifier predicted output and actual output.

  A. mean absolute error   B. root mean squared error

  C. mean squared error   D. mean relative error

V. Selecting data so as to assure that each class is properly represented in both the training and test set.

  A. cross validation   B. stratification

  C. verification   D. bootstrapping

VI. Bootstrapping allows us to

  A.   choose the same training instance several times.

  B.   choose the same test set instance several times.

  C.   build models with alternative subsets of the training data several times.

  D.   test a model with alternative subsets of the test data several times.

VII. For Bayes theorem to be applied, the following relationship between hypothesis H and evidence E must hold.

  A. $P(H|E) + P(H|\sim E) = 1$ B. $P(H|E) + P(\sim H|E) = 1$

  C. $P(H|E) + P(H|\sim E) = 0$ D. $P(H|E) + P(\sim H|E) = 0$

VIII. Which of the following problems is best solved using time-series analysis?

  A.   Predict whether someone is a likely candidate for having a stroke.

  B.   Determine if an individual should be given an unsecured loan.

  C.   Develop a profile of a star athlete.

D.   Determine the likelihood that someone will terminate their cell phone contract

IX. Which of the following would have a constant input in each epoch of training a Deep Learning model?

A. Weight between input and hidden layer

B. Weight between hidden and output layer

C. Biases of all hidden layer neurons

D. Activation function of output layer

X. Reinforcement learning: Immediate rewards

A. Must be positive

B. Can be used to estimate long-term rewards

C. Are problem-defined

D. Are stored in the Q table

SECTION- B

Section B contains five short answer questions with internal choice and all are compulsory. Each question carries eight marks.

UNIT-I

2. Determine which is the best approach for each of the problems I-VI.

   a. *supervised learning*

   b. *unsupervised clustering*

   c. *Simple data query*

   I.   What is the average weekly salary of all female employees under forty years of age?

   II. Develop a profile for credit card customers likely to carry an average monthly balance of more than $1000.00.

   III. Determine the characteristics of a successful used car salesperson.

   IV. What attribute similarities group customers holding one or several insurance policies?

   V.   Do single men play more golf than married men?

   VI. Determine whether a credit card transaction is valid or fraudulent.

OR

What is machine learning? Discuss various applications of machine learning techniques.

UNIT-II

3. What do you mean by Dimensionality reduction? Give an example of Dimensionality reduction.

OR

Which technique is used to arrange the given data items to similar groups? Explain with examples.

UNIT-III

4. Define various types of errors that may occur in a machine learning technique.

OR

What is a confusion matrix? What type of machine learning technique may be evaluated using confusion matrix?

UNIT-IV

5. Discuss various types of characteristics parameters for a time series.

OR

Using the following table answer questions (a) and (b)

a. Find the four - quarter centred moving average for 2001, quarter 3.

b. Find the four - quarter centred moving average for 2001, quarter 4

Year	Quarter	Y Sales value (£000)
2001	1	30
	2	20
	3	40
	4	50
2002	1	40
	2	30

UNIT-V

6. Differentiate the supervised and unsupervised pattern recognition.

OR

Discuss the applications of machine learning technique in Internet of Things (IoT).

SECTION -C

Section C contains 5 long answer questions any three are to be attempted. Each question carries 15 Marks

7. What are the differences between supervised and unsupervised machine learning. Discuss how to build and test a machine learning model.

8. Explain the importance and application of Principle component analysis (PCA)? Discuss with an example.

9. Differentiate between Bagging and Boosting ensemble techniques. Discuss their algorithms with their merits and demerits.

10. Explain Hidden Markov model (HMM) machine learning technique for sequential pattern recognition. Discuss a real life example where HMM may be applied.

11. What do you mean by Reinforcement machine learning? Write an algorithm for Q-learning. Explain the merits and demerits of Reinforcement learning.