Total No. of printed pages = 3

CSE 1817 OE 21

Roll No. of candidate

28/2/22-2021

B.Tech. 7th Semester End-Term Examination

CSE

MACHINE LEARNING

(New Regulation w.e.f 2017-18) &

(New Syllabus w.e.f 2018-19)

Full Marks – 70 Time – Three hours

The figures in the margin indicate full marks for the questions.

Answer Question No.1 and any *four* from the rest.

1. Answer the following : (MCQ/Fill in the blanks) (10 × 1 = 10)

(i) Real-Time decisions, AI game, learning tasks, skill acquisition, and robot navigation are applications of which of the following:

(a) Supervised Learning: Classification

(b) Reinforcement Learning

(c) Unsupervised Learning: Clustering

(d) Unsupervised Learning: Regression

(ii) What is perceptron?

(a) A single layer feed-forward neural network with pre-processing

(b) Neural network that contains feedback

(c) A double layer auto-associative neural network

(d) An auto-associative neural network

(iii) The Common classes of problems in machine learning is

(a) Classification (b) Clustering

(c) Regression (d) All of these

[Turn over

(iv) Machine learning algorithms build a model based on sample data known as _____

(a) Training data (b) Transfer data

(c) Validation data (d) Test data

(v) When performing regression or classification, which of the following is the correct way to preprocess the data

(a) Normalize the data → PCA → training

(b) PCA → normalize PCA output → training

(c) Normalize the data → PCA → normalize PCA output → training

(d) None of the above

(vi) Which of the following is a disadvantage of decision trees?

(a) Factor analysis

(b) Decision trees are robust to outliers

(c) Decision trees are prone to be overfit

(d) None of the above

(vii) 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)

(viii) If I am using all features of my dataset and I achieve 100% accuracy on my training set, but 70% on validation set, what should I look out for?

(a) Underfitting (b) Nothing, the model is perfect

(c) Overfitting (d) None of these

(ix) Choose the correct statement about the CART model:

(a) CART follows unsupervised learning techniques

(b) CART follows a supervised learning techniques

(c) CART adopt a greedy approach

(d) both (b) and (c)

(x) What are the different recommendation engine techniques:

(a) Content based filtering (b) Collaborative filtering

(c) Knowledge based system (d) All of these

2. (a) Define Machine Learning. Briefly explain the types of machine learning. (2+3=5)

(b) The values of the independent variable x and dependent variable y are given below.

X	Y
0	2
1	3
2	5
3	4
4	6

Find the least square regression line y=ax+b. Estimate the value of y when x=10. (10)

3. (a) How do we build a decision tree? (5)

(b) Explain the CART Algorithm. (10)

4. (a) What do you understand by model overfitting? What is bias and variance in Machine learning? (2+3=5)

(b) Explain with suitable example collaborative based recommendation system. (10)

5. (a) Formalize a mathematical model for a biological neuron. (7)

(b) A Perceptron can only converge on linearly separable data, support it with your answer. (8)

6. Write down the answers of the following questions: (3+3+9=15)

(a) What is the role of weights and bias in a neural network?

(b) What do you mean by exploding and vanishing gradients?

(c) Why does a Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) work better with image data and text data respectively?

7. (a) What is the importance of dimensionality reduction in machine learning? "Linear Discriminant Analysis often outperforms in a multi-class classification task when the class labels are known". Illustrate to support your answer. (2+7=9)

(b) What is K-means clustering? Give an examples of k-means clustering in real-life Data Applications. (3+3=6)