Total number of printed pages = 11

2022

MACHINE LEARNING

Full Marks – 100

Time – Three hours

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

Question 1 and 2 is compulsory, answer any *three* from the rest.

1. Multiple choice questions : 1×20=20

   (a) If machine learning model output involves target variable then that model is called as

   (i) descriptive model

   (ii) predictive model

   (iii) reinforcement learning

   (iv) All of the above

[Turn over

(b) In what type of learning labelled training data is used

    (i)   unsupervised learning

    (ii) supervised learning

    (iii) reinforcement learning

    (iv) active learning

(c) Which of the following is the best machine learning method ?

    (i)   scalable

    (ii) accuracy

    (iii) fast

    (iv) All of the above

(d) Data used to build a Machine Learning model.

    (i)   training data

    (ii) validation data

    (iii) test data

    (iv) hidden data

130/19/2nd Sem/PCSE 214   (2)

(e) The problem of finding hidden structure in unlabeled data is called

    (i)   supervised learning

    (ii) unsupervised learning

    (iii) reinforcement learning

    (iv) None of the above

(f) Of the following examples, which would you address using an supervised learning Algorithm ?

    (i)   Given email labelled as spam or not spam, learn a spam filter

    (ii) Given a set of news articles found on the web, group them into set of articles about the same story

    (iii) Given a database of customer data, automatically discover market segments and group customers into different market segments

    (iv) Find the patterns in market basket analysis

(g) You are given reviews of few netflix series marked as positive, negative and neutral. Classifying reviews of a new netflix series is an example of

    (i)   supervised learning

    (ii) unsupervised learning

    (iii) semisupervised learning

    (iv) reinforcement learning

(h) Which of the following is a good test dataset characteristic ?

    (i)   Large enough to yield meaningful results

    (ii) Is representative of the dataset as a whole

    (iii) Both (i) and (ii)

    (iv) None of the above

(i) Following is powerful distance metrics used by Geometric model

    (i)   Euclidean distance

    (ii) Manhattan distance

    (iii) Both (i) and (ii)

    (iv) Square distance

130/19/2nd Sem/PCSE 214   (4)

(j) The output of training process in machine learning is

    (i) machine learning model

    (ii) machine learning algorithm

    (iii) null

    (iv) accuracy

(k) PCA is

    (i) forward feature selection

    (ii) backword feature selection

    (iii) feature extraction

    (iv) All of the above

(l) Dimensionality reduction algorithms are one of the possible ways to reduce the computation time required to build a model.

    (i) True

    (ii) False

(m) Which of the following techniques would perform better for reducing dimensions of a data set ?

    (i) removing columns which have too many missing values

(ii) removing columns which have high variance in data

(iii) removing columns with dissimilar data trends

(iv) None of these

(n) Prediction is

    (i) the result of application of specific theory or rule in a specific case

    (ii) discipline in statistics used to find projections in multidimensional data

    (iii) value entered in database by expert

    (iv) independent of data

(o) Imagine a Newly-Born starts to learn walking. It will try to find a suitable policy to learn walking after repeated falling and getting up. Specify what type of machine learning is best suited ?

    (i) Classification

    (ii) Regression

    (iii) K-means algorithm

    (iv) Reinforcement learning

(p) Support Vector Machine is

   (i)   logical model

   (ii)   proababilistic model

   (iii) geometric model

   (iv) None of the above

(q) Impact of high variance on the training set ?

   (i)   Overfitting

   (ii)   Underfitting

   (iii) Both underfitting and overfitting

   (iv) Depents upon the dataset

(r) Which of the following are real world applications of the SVM ?

   (i)   Text and hypertext categorization

   (ii)   Image classification

   (iii) Clustering of news articles

   (iv) All of the above

(s) How can SVM be classified ?

    (i) It is a model trained using unsupervised learning. It can be used for classification and regression.

    (ii) It is a model trained using unsupervised learning. It can be used for classification but not for regression.

    (iii) It is a model trained using supervised learning. It can be used for classification and regression.

    (iv) It is a model trained using unsupervised learning. It can be used for classification but not for regression.

(t) Neural Networks are complex _____ with many parameters.

    (i) Linear Functions

    (ii) Nonlinear Functions

    (iii) Discrete Functions

    (iv) Exponential Functions

2. Write short notes on : 2×10=20

   (a) Machine Learning

   (b) Supervised Learning

   (c) Unsupervised Learning

   (d) Reinforcement Learning

   (e) Training and Test Data

   (f) Dependant and Independent Variables

   (g) Vector and Matrices in Training / Test Data

   (i) Feature Selection

   (j) One-hot-encoding

   (k) Binning

   (l) Normalization

   (m) Standardization

   (o) Logistic Regression.

3. (a) Write and explain different types of machine learning objectives. 10

(b) Make a prediction for the following sample dataset using Linear Regression machine learning technique. Use scatterplot to plot datapoint if necessary.   10

	(X)	(Y)	XY	X^2
1	1	3	3	1
2	2	4	8	4
3	1	2	2	1
4	4	7	28	16
5	3	5	15	9
Σ (Total)	11	21	56	31

4. (a) What is k-Nearest Neighbor algorithm ? Explain the kNN algorithm with an example.   10

(b) What is k-Means clustering algorithm? Explain k-Mean Clustering with an example.   10

5. (a) What are Bias and Variances ? Explian how biases and variances leads to underfitting and overfitting in Machine Learning.   10

(b) Explain Decision Tree Algorithm with an example.   10

6. (a) What are Support Vectors in SVM ? Explain SVM algorithm with an example.   10

   (b) What are Neurons ? Explain different types of Artificial Neural Network.   10

7. (a) What are RNN and CNN ? Explain their use in Machine Learning problems.   10

   (b) Design an Artificial Neural Network that can learn a function for the following data. Assume appropriate weights, activation functions and necessary number of neurons and layers as applicable.   10

X_1	X_2	Y
0	0	0
0	1	1
1	0	1
1	1	0