



HALDIA INSTITUTE OF TECHNOLOGY

(AN AUTONOMOUS INSTITUTION UNDER MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WEST BENGAL)

Paper Code: PCC-CS 603

Paper Name: Machine Learning

Time Allotted: 3 Hours

Full Marks: 70

The figures in the margin indicate full marks

Candidates are required to give their answers in their own words as far as practicable

Group – A

(Multiple Choice Type Questions)

Choose the correct alternatives from the followings:

15 x 1 = 15

1. (i) Which of the following is not a machine learning discipline? [CO1]

- a) Information theory b) Optimization control c) Physics d) Neuro statistics

(ii) Which of the following is not supervised learning? [CO1]

- a) PCA b) Naive Bayes c) Linear regression d) Decision Tree

(iii) Which supervised learning technique can process both numeric and categorical input attributes? [CO1]

- a) Bayes classifier b) KNN c) Logistic regression d) Linear regression

(iv) What is the minimum number of features required to perform clustering? [CO5]

- a) 0 b) 1 c) 2 d) 3

(v) PCA is [CO1]

- a) Forward feature selection b) Backward feature selection
c) Feature extraction d) None of the above

(vi) Supervised learning and unsupervised clustering both require which is correct according to the statement. [CO1]

- a) Input attribute b) Hidden attribute c) Output attribute d) Categorical output

(vii) Stacking builds ensembles in _____. [CO1]

- a) Series. b) Parallel. c) Series and parallel. d) None of the above

(viii) What is true about an ensemble classifier? [CO4]

1. Classifiers that are more “sure” can vote with more conviction
2. Classifiers can be more “sure” about a particular part of the space
3. Most of the time, it performs better than a single classifier
a) 1 and 2 b) 1 and 3 c) 2 and 3 d) All the above

(ix) Bagging comes under which _____ method? [CO1]

- a) Sequential methods. b) Parallel method. c) Line method. d) Dot method.

(x) _____ uses ensembles of ML models each capturing a specific subspace of predictor space. [CO4]

- a) Bagging. b) Boosting. c) Stacking. d) None of the above

(xi) Which of the following is not an example of an ensemble method? [CO1]

- a) AdaBoost b) Decision tree c) Random Forest d) Bootstrapping

(xii) The systematic components of time series which follow a regular pattern of variations are called: [CO1]

- a) Signal b) Noise c) Additive model d) Multiplicative model

(xiii) CNN is mostly used when there is an? [CO1]

- a) unstructured data b) structured data c) None of the above d) All of above

(xiv) Which of the following is a subset of machine learning? [CO1]

- a) Numpy b) Deep Learning c) Scipy d) All of the above

(xv) Which of the following is/are the Limitations of Deep Learning? [CO3]

- a) Data Labeling b) Obtain Huge Training Datasets
c) Both A and B d) None of the above

Group – B

(Short Answer Type Questions)

Attempt any three from the followings:

3 x 5 = 15

2. How is supervised learning different from unsupervised learning? What is the cost function in machine learning? [CO1, CO2] (2+3)

3. i) Explain briefly the working principles of the k-means algorithm. [CO1, CO2]

ii) What are the differences between K-means clustering and K-nearest neighbor(KNN)? [CO1, CO2, CO3] (3+2)

4. What are the advantages and disadvantages of the Random Forest algorithm? [CO3] (5)

5. Explain cyclic variations. [CO1, CO4] (5)

6.i) What is an activation function in an artificial neuron? Give some examples. [CO1]

ii) Define a perceptron. [CO1] (2+1+2)

Group – C

(Long Answer Type Questions)

Attempt any four from the followings:

4 x 10 = 40

7. (i) Explain the working principle of the support vector machine algorithm. [CO1]

(ii) Write the difference between linear and non-linear SVM. [CO1]

(iii) Find the optimal hyperplane for the set of the following data points. (1,1),(2,1),(1,1),(2,1),(4,0),(5,1),(5,-1),(6,0) and classes

Class 1: {(1,1),(2,1),(1,-1),(2,-1)}

Class 2: {(4,0),(5,1),(5,-1),(6,0)} [CO1, CO5, CO6] (3+1+6)

8. (i) What are the common problems with sparse datasets in machine learning? [CO1, CO2]

(ii) Write a short note on matrix factorization. [CO1, CO2] (5+5)

9. a) What is the principle behind the Random Forest ensemble model? [CO3]

b) Explain the differences between Random Forest and Decision Tree? [CO3, CO4] (5+5)

10. Define Time series and time series data. [CO1]

What are the different methods for time series forecasting? [CO1, CO4]

Explain with examples the different components of time series. [CO1, CO4] (2+3+5)

11. i) Explain Bayesian Inference with a suitable example. [CO1]

ii) What is a graphical model in machine learning? What are common graphical models in machine learning? Mention application areas where the graphical model can be used. [CO1] 5+(2+2+1)

12. i) Discuss the working strategy of semi-supervised learning. [CO1]

ii) State the difference between Semi-supervised and Reinforcement learning. [CO1,CO3,CO4]

iii) What are the limitations of semi-supervised learning? [CO1,CO3]

(5+3+2)