



HALDIA INSTITUTE OF TECHNOLOGY

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

Paper Code: PCC-CS 503

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) A software designed to evaluate whether a customer account is hacked or not is an example of

a) Unsupervised learning b) supervised classification [CO1]

c) Supervised regression d) None of these

(ii) How many coefficients do you need to estimate a simple regression model (one independent variable)?

a) 1 b) 2 c) 3 d) 4 [CO1]

(iii) Which of the following is a measure of the spread of data?

a) Mean b) Variance c) Covariance d) All of the above [CO1]

(iv) Which of the following is an alternative technique to principal component analysis?

a) Factor analysis b) Independent component analysis

c) Latent semantic analysis d) All of the above [CO1]

(v) Which supervised learning technique can process both numeric and categorical input attributes?

a) Bayes classifier b) Linear regression c) Logistic regression d) None of the above [CO1]

(vi) Naïve Bayes Algorithm belongs to which category?

a) Supervised b) Reinforcement c) Unsupervised d) None of these [CO2]

(vii) __ is the machine learning algorithm that can be used with unlabeled data.

a) Regression algorithm b) Clustering algorithm [CO2]

c) Instance-based algorithm d) All of the above

(viii) What does K stand for in K mean algorithm?

a) Number of clusters b) Number of data c) Number of attributes d) Number of iterations [CO2]

(ix) Boosting builds ensembles in _____.

a) Series b) Parallel c) Series and parallel d) None of the above [CO3]

(x) Bagging comes under which _____ method?

a) Sequential methods b) Parallel method c) Line method d) Dot method [CO3]

(xi) An orderly set of data arranged in accordance with their time of occurrence is called: [CO4]

a) Arithmetic series b) Harmonic series c) Geometric series d) Time series

(xii) Secular trend can be measured by:

[CO4]

a) Two methods b) Three methods c) Four methods d) Five methods

(xiii) The unsystematic sequence which follows irregular pattern of variations is called: [CO4]

- a) Noise b) Signal c) Linear d) Non-linear

(xiv) The general limitations of back propagation rule is/are

- a) Scaling b) Slow convergence c) Local minima problem d) All of the above

(xv) How many types of feedback does reinforcement provide?

[CO5], [CO6]

- a) 1 b) 2 c) 3 d) 4

Group – B

(Short Answer Type Questions)

Attempt any three from the followings:

3 x 5 = 15

2. How logistic regression is different compared to linear regression? Explain with suitable example. [CO1] 5

3. (i) What is the Euclidian distance between two data points: A(1,3) and B(2,3)? [CO1]

(ii) Which machine learning algorithm is called lazy learner and why? [CO1] 2+3

4. (i) What do you mean by clustering?

[CO2]

(ii) Write note on the different types of clustering. [CO2] 1+4

5. (i) Define seasonal index.

[CO4]

(ii) Explain cyclic variations.

[CO4]

2+3

6. (i) State the need of scaling in machine learning.

[CO5], [CO6]

(ii) Name some scaling techniques in ML with example

[CO5], [CO6]

2+3

Group – C

(Long Answer Type Questions)

Attempt any four from the followings:

4 x 10 = 40

7. (i) The sales of a company (in million dollars) for each year are shown in the table bellow

[CO1]

S. No.	Year	Sales
1	2005	12
2	2006	19
3	2007	29
4	2008	37
5	2009	45

Find the least square regression line, $y = ax + b$.

(ii) Use the least square regression line as a model to estimate the sales of the company in 2012. [CO1] 5+5

8. (i) Explain matrix factorization and where it is used? [CO2]

(ii) What are the advantages and disadvantages of PCA algorithm? [CO2] 5+5

9. (i) What is the general idea of boosting technique? [CO3]

(ii) Compare bagging and boosting. [CO3] 5+5

10. (i) Explain the method of fitting a straight line. [CO4]

(ii) State the two normal equations used in fitting a straight line. [CO4] 5+5

11. (i) Applying the k-means algorithm, find two clusters in the following data. [CO4]

x	185	170	168	179	182	188
y	72	56	60	68	72	77

10

12. (i) Explain the architecture of an artificial neural network (ANN).

[CO5], [CO6]

(ii) What is the role of the Activation functions in Neural Networks?

[CO5], [CO6]

5+5