

(6 Pages)

Reg. No. :

Code No. : 10468 E Sub. Code : CACS 41

B.Sc. (CBCS) DEGREE EXAMINATION, APRIL 2023.

Fourth Semester

Computer Science — Allied

MACHINE LEARNING TECHNIQUES

(For those who joined in July 2021 onwards)

Time : Three hours

Maximum : 75 marks

PART A — (10 × 1 = 10 marks)

Answer ALL questions.

Choose the correct answer :

1. Machine learning is an application of _____
 - (a) Block chain
 - (b) Artificial intelligence
 - (c) Both (a) and (b)
 - (d) None of these

2. The categories in which machine learning approaches can be traditionally categorized are _____
 - (a) Supervised learning
 - (b) Unsupervised learning
 - (c) Reinforcement learning
 - (d) All of the above
3. Logistic regression is a _____ regression technique that is used to model data having a binary outcome.
 - (a) Linear
 - (b) Nonlinear
 - (c) Numeric
 - (d) Nonnumeric
4. _____ is a disadvantage of decision trees.
 - (a) Decision trees are robust to outlier
 - (b) Decision trees are prone to be overfit
 - (c) Both (a) and (b)
 - (d) None of these
5. Scikit-learn depends on _____ and _____ python packages.
 - (a) NumPy and SciPy
 - (b) NumPy and StrPy
 - (c) NicPy and SciPy
 - (d) None of these



6. The _____ notebook is an interactive environment for running code in the browser.
- (a) Jupyter (b) Jnode
(c) Jsnode (d) Kupyter
7. Which is needed by K-means clustering?
- (a) defined distance metric
(b) number of clusters
(c) initial guess as to cluster centroids
(d) all of these
8. Which of the following clustering requires merging approach?
- (a) Partitional
(b) Hierarchical
(c) Naive Bayes
(d) None of the mentioned
9. The subfield of data science/machine learning related to text is called _____
- (a) Image processing
(b) Regression
(c) Classification
(d) Natural language processing

10. TF-IDF stands for
- (a) Text Frequency - Inverse Document Frequency
(b) Term Frequency - Index Document Frequency
(c) Term Frequency - Inverse Document Frequency
(d) All of these

PART B — (5 × 5 = 25 marks)

Answer ALL questions, choosing either (a) or (b).
Each answer should not exceed 250 words.

11. (a) Why Python used in machine learning?
- Or
- (b) What is data exploration in data visualization?
12. (a) Define simple linear regression.
- Or
- (b) What is credit classification? Explain.
13. (a) Discuss about matrix factorization.
- Or
- (b) List the Scikit - Learn library for machine learning.



14. (a) How does Clustering works?

Or

(b) Write K-means algorithm.

15. (a) Explain about sentiment classification.

Or

(b) Discuss about the challenges in text analysis.

PART C — (5 × 8 = 40 marks)

Answer ALL questions, choosing either (a) or (b).

Each answer should not exceed 600 words.

16. (a) Write about framework for developing machine learning models.

Or

(b) What is the advantages of machine learning?

17. (a) What are the steps in building a regression? Explain.

Or

(b) Explain in detail about multiple linear regression.

18. (a) Write and explain Gradient r Algorithm.

Or

(b) Why we need advanced regression model?

19. (a) What are the advantages of hierarchical clustering algorithms?

Or

(b) Illustrate advanced machine learning algorithm.

20. (a) Explain Naive-Bayes model for sentiment classification.

Or

(b) Discuss text analysis with Tf-IDF vectorization.

