

(6 Pages)

Reg. No. :

Code No. : 6071

Sub. Code : BCAM 22/
ZCAM 22

M.C.A. (CBCS) DEGREE EXAMINATION, APRIL 2022

Second Semester

Computer Applications — Core

MACHINE LEARNING USING PYTHON

(For those who joined in July 2020-2021 onwards)

Time : Three hours

Maximum : 75 marks

PART A — (10 × 1 = 10 marks)

Answer ALL questions.

Choose the correct answer :

1. ——— algorithms are used when the outputs may have any numerical value within a range.
- (a) KNN (b) Regression
(c) Clustering (d) Classification

2. ——— algorithms learn a function that can be used to predict the output associated with new inputs.

(a) Reinforcement (b) Un-supervised
(c) Supervised (d) Semi-Supervised

3. Among the following identify the one in which dimensionality reduction reduces ———

(a) performance (b) entropy
(c) stochastic (d) collinearity

4. Which of the following are common classes of problems in machine learning? ———

(a) Regression (b) Classification
(c) Clustering (d) All the above

5. Which is the disadvantages of decision trees?
—————

(a) Factor analysis
(b) Decision trees are robust to outliers
(c) Decision trees are prove to be over fit
(d) None



6. ——— transformation changes linear relationships between variables, and thus changes the correlation, between variables.

- (a) Linear (b) Non linear
- (c) Both (a) and (b) (d) None

7. Types of cross validation ———

- (a) Leave one out, hold out
- (b) K-fold, nested
- (c) Time series, leave port
- (d) All the above

8. In ——— cross validation, the original dataset is equally partitioned into K-subparts of folds.

- (a) K-fold (b) Nested
- (c) Holdout (d) None

9. A ——— is a way to coding and automate the workflow it takes to produce a machine learning model.

- (a) Pipeline (b) Binning
- (c) Both (a) and (b) (d) None

Page 3 Code No. : 6071

10. Challenges associated with machine learning pipeline ———

- (a) Data collection and cleaning
- (b) Visualisation
- (c) Model validation
- (d) All the above

PART B — (5 × 5 = 25 marks)

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

11. (a) Why do we need python is machine language?

Or

(b) What is supervised learning?

12. (a) List out the challenges in unsupervised learning.

Or

(b) Differentiate : training data versus test data.

13. (a) How to utilize expert knowledge?

Or

(b) What do you mean one-hot-encoding in categorical variables?

Page 4 Code No. : 6071

[P.T.O]



14. (a) What is grid search?

Or

(b) Give a brief note on the danger of over fitting the parameters and the validation set.

15. (a) How to avoid redundant computation?

Or

(b) Write short note on parameter selection with preprocessing.

PART C — ($5 \times 8 = 40$ marks)

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

16. (a) What do you mean by classification and regression? Explain the term generalization, overfitting and underfitting.

Or

(b) Write short note on Naive Bayes classifiers and SVM.

Page 5

Code No. : 6071

17. (a) Explain dimensionality reduction, feature extraction and manifold learning.

Or

(b) Write short note on the following :

(i) Preprocessing

(ii) Agglomerative clustering

(iii) DBSCAN.

18. (a) What do you mean by univariate non linear transformation? Explain it in detail.

Or

(b) Discuss in detail about interactions and polynomials.

19. (a) What is cross-validation? Explain its types.

Or

(b) Explain evaluation matrix and scanning in detail.

20. (a) How to build pipelines? Give a brief note on using pipelines in grid searches.

Or

(b) Describe the general pipeline interface in detail.

Page 6

Code No. : 6071

