

B.C.A. (CBCS) DEGREE EXAMINATION,
NOVEMBER 2023.

Fifth Semester

Computer Application – Core

MACHINE LEARNING WITH PYTHON

(For those who joined in July 2021-2022)

Time : Three hours

Maximum : 75 marks

PART A — (10 × 1 = 10 marks)

Answer ALL questions.

Choose the correct answer:

- _____ language is mostly used for machine learning applications.
 - (a) C
 - (b) C++
 - (c) Java
 - (d) Python

- _____ is a GUI toolkit for Python.
 - (a) Tkinter
 - (b) Kotlin
 - (c) Tlib
 - (d) Widgets
- A tree structure modeling the data with piecewise constant segments is known as a
 - (a) regression tree
 - (b) CART
 - (c) pruning
 - (d) linear segments
- _____ means "from before".
 - (a) Association analysis
 - (b) FP-growth
 - (c) Apriori
 - (d) FP-Trees
- Hadoop stores the huge amount of data through a system called
 - (a) HDFS
 - (b) HDDS
 - (c) GFS
 - (d) DSA
- Name the programming model used for efficient processing in parallel over large data-sets in a distributed manner.
 - (a) Hadoop
 - (b) GoogleMap
 - (c) AWs
 - (d) MapReduce

- Matplotlib has a great tool, called _____, that can add text near data in a plot.
 - (a) Splitting
 - (b) Merging
 - (c) Annotations
 - (d) Drawing
- In Exact Bayes calculation is limited to the two matching with there
 - (a) Diagnosis value
 - (b) Probabilistic condition
 - (c) Characteristics
 - (d) Clustering
- Bayes' rule tells us how to _____ the symbols in a conditional probability statement.
 - (a) assign
 - (b) swap
 - (c) remove
 - (d) duplicate
- _____ regression uses the identity matrix multiplied by some constant λ .
 - (a) linear
 - (b) ridge
 - (c) non-linear
 - (d) regular

PART B — (5 × 5 = 25 marks)

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

Each answer should not exceed 250 words.

- (a) What is meant by unsupervised learning?
Or
(b) State the significance of NumPy library.
- (a) How can we measure the line that best separates the data?
Or
(b) List the general approach to logistic regression.
- (a) Comment on Ridge regression.
Or
(b) Explain the features of Model tree.
- (a) Summarize the Unsupervised learning technique.
Or
(b) Narrate the property of association rule mining.

15. (a) Demonstrate the applications of the SVD.

Or

(b) Elaborate the key points about MapReduce.

PART C — (5 × 8 = 40 marks)

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

Each answer should not exceed 600 words.

16. (a) Give a pseudo code for kNN algorithm and explain.

Or

(b) What is the purpose of decision tree? How to construct decision tree?

17. (a) Write note on AdaBoost algorithm.

Or

(b) How to find the maximum margin using SVM? Explain.

18. (a) Discuss the bias/variance tradeoff.

Or

(b) Examine the features of Tree pruning.

Page 5 Code No. : 20508 E

19. (a) Explain about association analysis.

Or

(b) Elaborate the FP-growth algorithm.

20. (a) Describe the features of Hadoop Streaming.

Or

(b) Enumerate the services available on AWS.

Page 6 Code No. : 20508 E