S.No. : 576 NAI4401

No. of Printed Pages : 04

Following Paper ID and Roll No. to be filled in your Answer Book.

PAPER ID : 33902 Roll No. 1 2 3 0 4 3

B. Tech Examination 2024-25

(Even Semester)

CONCEPTS OF MACHINE LEARNING WITH PYTHON

Time : Three Hours] *[Maximum Marks : 60*

Note :– Attempt all questions.

SECTION–A

1. Attempt all parts of the following: $1 \times 8 = 8$

   (a) Define the term "version space" in machine learning.

   (b) What is the role of the scikit-learn library in python for machine learning?

   (c) What is the difference between linear regression and logistic regression?

[P. T. O.

(d) Write a python code snippet to split a dataset into training and testing sets using train_test_split.

(e) Explain the objective of K-means clustering in unsupervised learning.

(f) Write a python code snippet to perform principal component analysis(PCA) using scikit-learn.

(g) What is the purpose of the naive Bayes classifier in probabilistic learning?

(h) Define the term "boosting" in ensemble learning.

SECTION – B

2. Attempt any two parts of the following: $2 \times 6 = 12$

(a) Explain the concept of probably approximately correct (PAC) learning. How does it ensure model generalization?

(b) Describe the working of the decision tree algorithm(ID3). Provide a python code snippet to implement a decision tree classifier using scikit-learn.

(c) Discuss the process of K-means clustering. How does it differ from hierarchical clustering?

(d) Explain the Adaboost algorithm in ensemble learning. How does it improve classification performance?

SECTION–C

Note:- Attempt all questions. Attempt any two parts from each question. $5 \times 8 = 40$

3. (a) Discuss the role of hypthesis space and VC dimension in machine learning. How do they impact model complexity?

(b) Explain the steps to design and evaluate a machine learning experiment. Provide a python code snippet to perform k-fold cross-validation using scikit-learn.

(c) Evaluate the challenges of handling imbalanced datasets in machine learning. Suggest two techniques to address this issue with examples.

4. (a) Describe the architecture and working of support vector machines(SVM) performance?

(b) Write a python code snippet to implement logistic regression using scikit-learn. Include steps for model training, prediction, and evaluation using accuracy and ROC curve.

(c) Compare and contrast the Random Forest and CART algorithms. Evaluate their suitability for large datasets.

5. (a) Explain the expectation maximization(EM) algorithm for Gaussian Mixture Models(GMM). Discuss its application in clustering with an example.

(b) Write a python code snippet to implement K-means clustering using scikit-learn. Explain how the algorithm optimizes cluster assignments.

(c) Evaluate the strengths and limitations of principal component analysis(PCA) for dimensionality reduction. Provide an example of its use in a real-world scenario.

6. (a) Discuss the Naive Bayes classifier, including its assumptions and applications. Provide an example of its use in spam email detection.

(b) Write a python code snippet to implement a Random Forest classifier using scikit-learn. Include steps for hyperparameter tuning using grid search.

(c) Compare and contrast the adaboost and bagging ensemble methods. Evaluate their effectiveness for improving model performance in classification tasks.

⌘⌘⌘