

AIE111 Artificial Intelligence

25/12/2024

Sheet 2

Question 1: Loading and Exploring the Wine Dataset

- A.** Write Python code to load the Wine dataset using `sklearn.datasets.load_wine`.
- B.** Display the first 5 rows of the dataset along with their corresponding labels.

Question 2: Splitting the Dataset

Write code to split the Wine dataset into training and testing sets using an 80:20 ratio.

Question 3: Implementing Naive Bayes

- A.** Train a Naive Bayes classifier on the Wine dataset using the `GaussianNB` class from `sklearn`.
- B.** Evaluate the accuracy of the classifier on the test set.

Question 4: Decision Tree Implementation

- A.** Train a Decision Tree classifier on the Wine dataset using the `DecisionTreeClassifier` class from `sklearn`.
- B.** Visualize the Decision Tree using `plot_tree`.
- C.** Evaluate the accuracy of the Decision Tree on the test set.

Question 6: Feature Importance

- A.** Identify the most important features for classification using the Decision Tree model.
- B.** Plot a bar chart of feature importance scores.

Question 7: Classification Report

Write code to generate a classification report for both Naive Bayes and Decision Tree models using `classification_report` from `sklearn.metrics`.

Question 8: Confusion Matrix Analysis

- A. Write code to compute the confusion matrix for both Naive Bayes and Decision Tree classifiers using the test set.
- B. Visualize the confusion matrix using a heatmap from the `seaborn` library.
- C. Explain the meaning of True Positives, False Positives, True Negatives, and False Negatives in the context of the confusion matrix.

Question 9: Comparison of Algorithms

Write a function to compare the precision, recall, and F1 scores of Naive Bayes and Decision Tree classifiers on the Wine dataset.

Question 10: Alternative Dataset: Breast Cancer

- A. Load the Breast Cancer dataset from `sklearn.datasets`.
- B. Train both Naive Bayes and Decision Tree classifiers on this dataset.
- C. Evaluate and compare their performances in terms of accuracy, precision, and recall.

Question 11: Naive Bayes Probabilities

Given a dataset with the following class distributions:

Feature Class A Class B

X1 = 1 0.6 0.4

X2 = 2 0.7 0.3

A. Calculate the posterior probability of Class A and Class B for a sample with $X_1 = 1$ and $X_2 = 2$ using Python. Assume equal priors.

Question 12: Entropy and Information Gain

A. Write Python code to calculate the entropy of a dataset with two classes: 30 instances of Class A and 20 instances of Class B.

B. Write code to compute the information gain for splitting the dataset on a feature that separates it into subsets: 20 Class A, 5 Class B (subset 1) and 10 Class A, 15 Class B (subset 2).

Question 13: Gini Impurity

A. Write Python code to calculate the Gini Impurity for a dataset with the following distribution: 50% Class A, 30% Class B, and 20% Class C.

Question 14: Decision Tree on Digits Dataset

A. Load the Digits dataset from `sklearn.datasets`.

B. Train a Decision Tree classifier on this dataset and report its accuracy.

C. Visualize the Decision Tree using `plot_tree`.

Question 15: Comparing Classifiers on Custom Dataset

A. Load a custom dataset of your choice (CSV file with labeled data).

B. Train both Naive Bayes and Decision Tree classifiers on this dataset.

C. Compare their performance in terms of accuracy, precision, and recall.