Lab-Experimet-3 and 4

Python Lab: Comparing Logistic Regression and Naïve Bayes for Classification

Explore the application of Logistic Regression and Naïve Bayes classifiers on a publicly available dataset. Your task is to compare the performance of these classifiers using essential performance metrics such as accuracy, precision, recall, F1-score, and the confusion matrix.

Dataset Description:

You will use the famous Iris dataset, which contains information about iris flowers, including sepal and petal dimensions. It consists of 150 samples, with each sample belonging to one of three classes: Iris Setosa, Iris Versicolor, and Iris Virginica.

Tasks:

1. Load the Iris dataset using scikit-learn or any other preferred method.

3. Draw the required plots to visualize the dataset

4. Split the dataset into training and testing sets (80% training, 20% testing).

5. Implement Logistic Regression and Naïve Bayes classifiers using scikit-learn.

6. Train the classifiers on the training data.

7. Evaluate the performance of each classifier using accuracy, precision, recall, and F1-score.

8. Draw the AUC and RUC curve for each algorithm

9. Draw the confusion matrix for each classifier.

10. Compare the performance of the two classifiers and discuss your findings.

Instructions:

- Use Python and scikit-learn library to complete the tasks.

- Write well-commented code for each task.

- Explain the choice of evaluation metrics and any observations made during analysis.

- Include appropriate visualizations and/or tables to support your analysis.

- Write a conclusion summarizing your findings.

Marks Distribution:

1. Data Loading and Preprocessing: 5 marks

2. Model Implementation: 10 marks

3. Model Evaluation (including all required performance metrics and confusion matrix): 20 marks

4. Comparison and Analysis: 10 marks

5. Clarity of Code and Comments: 5 marks

Total: 50 marks

Note: You can refer to the scikit-learn documentation for details on loading the Iris dataset, implementing classifiers, and drawing confusion matrices. Ensure to provide clear explanations and interpretations of the results obtained.