Assignment: 2

Submission date: 25 October 2024

Title: Comparative Analysis of Random Forest, Decision Tree, and SVM Classifiers

Objective: Compare the performance of Random Forest, Decision Tree, and SVM classifiers on a given dataset.

Dataset: Use the Iris dataset (available in scikit-learn library) or any other dataset of your choice.

Tasks:

Part 1: Data Preprocessing (10 points)

1. Load the dataset and explore its features.
2. Handle missing values (if any).
3. Normalize/scale the data (if necessary).

Part 2: Decision Tree Classifier (20 points)

1. Train a Decision Tree classifier on the preprocessed data.
2. Tune hyperparameters using GridSearchCV.
3. Evaluate the model's performance using accuracy, precision, recall, and F1-score.
4. Visualize the decision tree using plot_tree.

Part 3: Random Forest Classifier (20 points)

1. Train a Random Forest classifier on the preprocessed data.
2. Tune hyperparameters using GridSearchCV.
3. Evaluate the model's performance using accuracy, precision, recall, and F1-score.
4. Analyze feature importance using feature_importances_.

Part 4: SVM Classifier (20 points)

1. Train an SVM classifier on the preprocessed data.
2. Tune hyperparameters using GridSearchCV.
3. Evaluate the model's performance using accuracy, precision, recall, and F1-score.
4. Visualize the decision boundary using plot_decision_boundary.

Part 5: Comparison and Conclusion (30 points)

1. Compare the performance of the three classifiers.
2. Discuss the strengths and weaknesses of each classifier.
3. Provide recommendations for future improvements.

Deliverables:

1. Code (Python notebook or script)
2. Report (PDF or Word document)

Grading Criteria:

1. Code quality and organization (20 points)
2. Accuracy and completeness of results (30 points)
3. Clarity and coherence of report (20 points)
4. Adherence to instructions and formatting (30 points)

Submission Guidelines:

1. Submit the code and report separately.
2. Use a clear and concise filename format (e.g., "Random_Forest_Assignment.ipynb" and "Random_Forest_Report.pdf").