Flower Classification Report

Github : https://github.com/NovasusTV/Dorsret

Introduction

This report focuses on the classification of flowers based on their features using the Iris Flowers dataset. The dataset consists of measurements of petal length, petal width, sepal length, and sepal width for 150 iris flowers belonging to three different species: setosa, versicolor, and virginica.

Exploratory Data Analysis

We began by performing exploratory data analysis on the dataset. This step involved understanding the dataset's structure, checking for missing values, and gaining insights into the distribution of the features. We also visualized the data using pair plots to observe relationships between different features and the flower species.

Data Preprocessing

To prepare the data for classification, we split it into input features (X) and target variable (y). We then split the data into training and testing sets using the train_test_split function from scikit-learn. This allowed us to train the model on a portion of the data and evaluate its performance on unseen examples.

Support Vector Machines (SVM)

Support Vector Machines (SVM) is a powerful algorithm used for classification tasks. In this project, we employed the SVM algorithm to classify the flowers based on their features. Specifically, we used the SVC (Support Vector Classifier) class from scikit-learn to train an SVM model.

Model Training and Evaluation

We trained the SVM model on the training data using the fit method. This process involved finding an optimal hyperplane that maximally separates the different flower species in the feature space. After training the model, we made predictions on the test set using the predict method and calculated the accuracy score to evaluate its performance.

Hyperparameter Tuning

To further improve the model's performance, we conducted hyperparameter tuning. We utilized the GridSearchCV class from scikit-learn to perform grid search with cross-validation. Grid search involved systematically evaluating the model's performance across different combinations of hyperparameters. By finding the best hyperparameters, we enhanced the model's accuracy and generalization ability.

Model Evaluation and Visualization

We evaluated the final SVM model on the test set and calculated the accuracy score as well as generated a classification report. The classification report provided metrics such as precision, recall, and F1-score for each flower species, giving insights into the model's performance.

Visualizations played a crucial role in understanding the model's behavior. We plotted training loss and accuracy to observe the model's learning progress. Additionally, we used a confusion matrix to visualize the model's classification results, providing a comprehensive overview of its performance for each flower species.

Conclusion

In conclusion, we successfully employed Support Vector Machines (SVM) to classify flowers based on their features. Through exploratory data analysis, data preprocessing, model training, hyperparameter tuning, and evaluation, we achieved accurate classification results. The SVM algorithm, coupled with careful hyperparameter selection, proved to be effective in accurately distinguishing between different flower species.

The findings from this project demonstrate the importance of data exploration, preprocessing, and model evaluation in achieving reliable and robust machine learning models. The insights gained from this flower classification task can have practical applications in various domains.