

Documentation: Optimization of Gradient Boosting Classifier Parameters for Wine Dataset using GridSearchCV

Introduction

This Python script demonstrates the process of optimizing a Gradient Boosting Classifier using the GridSearchCV function from Scikit-learn library on the Wine dataset. The main aim is to find the best hyperparameters that would give the most accurate predictions.

Dependencies

This script uses the following Python libraries:

- `time`: To calculate the execution time of the GridSearchCV function.
- `warnings`: To ignore any warnings generated during the execution of the script.
- `numpy`: For numerical computations.
- `sklearn.datasets`: To load the wine dataset.
- `sklearn.model_selection`: Contains the `train_test_split` function to split the dataset into training and testing sets and `GridSearchCV` for hyperparameter tuning.
- `sklearn.ensemble`: Contains the `GradientBoostingClassifier`, the machine learning model used in the script.
- `sklearn.metrics`: To generate the `classification_report` and `confusion_matrix` for model evaluation.
- `matplotlib.pyplot` and `seaborn`: For creating visualizations.

The Script

Function Definition: `optimize_model`

This function performs hyperparameter tuning on a given model using GridSearchCV, trains the model on the best parameters, makes predictions on the test set, and prints the best parameters and model evaluation metrics.

The function takes as input a model, a dictionary of parameters for GridSearchCV, and the dataset (features `x` and labels `y`).

The function performs the following steps:

1. Data Split: It uses `train_test_split` from `sklearn.model_selection` to split the dataset into training and testing sets.
2. Grid Search: It initializes a GridSearchCV object with the input model and parameters, and fits this object to the training data.
3. Best Model Training and Prediction: It retrieves the best model from the grid search, uses it to make predictions on the test set, and prints the best parameters and their score.
4. Evaluation: It prints the classification report and displays a confusion matrix and a feature importance plot.

Main Part of the Script

The main part of the script does the following:

1. Loading the Dataset: It loads the Wine dataset from `sklearn.datasets`.
2. Defining the Model and Parameters: It initializes a Gradient Boosting Classifier and defines a dictionary of parameters for the grid search.
3. Calling the `optimize_model` Function: It calls the `optimize_model` function with the model, parameters, and dataset as arguments.