base.py

Summary

BaseAutoCV is an abstract base class for implementing automated cross-validation and hyperparameter optimization using Bayesian optimization.

Dependencies

Standard Library

- abc
- typing

Other

- numpy
- pandas
- optuna
- pydantic
- sklearn

Description

The `BaseAutoCV` class is an abstract base class that provides a framework for implementing automated cross-validation and hyperparameter optimization using Bayesian optimization. It inherits from both `ABC` (Abstract Base Class) and `BaseEstimator` from scikit-learn.

The class is designed to be subclassed and not instantiated directly. It provides a structure for creating estimators that can automatically find optimal hyperparameters using Optuna, a hyperparameter optimization framework. The class includes methods for prediction, explanation, and optimization, as well as properties for accessing the optimization study and cross-validation results.

The main functionality is implemented in the `optimize` method, which uses Optuna to perform Bayesian optimization of the model hyperparameters. It creates a study object with TPESampler and HyperbandPruner, then optimizes an objective function that performs cross-validation for each trial. The best parameters found during the optimization process are stored and used to fit the final model.

The class also includes utility methods for validating input data (`_validate_X` and `_validate_y`) and abstract methods that need to be implemented by subclasses (`predict`, `predict_proba`, `scorer`, and `explain`). These methods ensure proper data handling and provide a consistent interface for different types of models.

The implementation includes Pydantic validation for input parameters and type hints for better code safety.