Machine Learning

Session 18 - PL

Ensemble Learning and Hyperparameter Optimization

Degree in Applied Data Science

2024/2025

Ensemble Learning with Scikit-Learn

- Bagging:

- https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.BaggingClassifier.html


```
sklearn.ensemble.BaggingClassifier

class sklearn.ensemble.BaggingClassifier(estimator=None, n_estimators=10, *, max_samples=1.0, max_features=1.0,
bootstrap=True, bootstrap_features=False, oob_score=False, warm_start=False, n_jobs=None, random_state=None, verbose=0)
                                                                                                              [source]
```

- https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.BaggingRegressor.html


```
sklearn.ensemble.BaggingRegressor

class sklearn.ensemble.BaggingRegressor(estimator=None, n_estimators=10, *, max_samples=1.0, max_features=1.0,
bootstrap=True, bootstrap_features=False, oob_score=False, warm_start=False, n_jobs=None, random_state=None, verbose=0)
                                                                                                              [source]
```

Ensemble Learning with Scikit-Learn

- Boosting:

  - https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html


```
sklearn.ensemble.GradientBoostingClassifier

class sklearn.ensemble.GradientBoostingClassifier(*, loss='log_loss', learning_rate=0.1, n_estimators=100,
subsample=1.0, criterion='friedman_mse', min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0,
max_depth=3, min_impurity_decrease=0.0, init=None, random_state=None, max_features=None, verbose=0,
max_leaf_nodes=None, warm_start=False, validation_fraction=0.1, n_iter_no_change=None, tol=0.0001, ccp_alpha=0.0)[source]
```

  - https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingRegressor.html


```
sklearn.ensemble.GradientBoostingRegressor

class sklearn.ensemble.GradientBoostingRegressor(*, loss='squared_error', learning_rate=0.1, n_estimators=100,
subsample=1.0, criterion='friedman_mse', min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0,
max_depth=3, min_impurity_decrease=0.0, init=None, random_state=None, max_features=None, alpha=0.9, verbose=0,
max_leaf_nodes=None, warm_start=False, validation_fraction=0.1, n_iter_no_change=None, tol=0.0001, ccp_alpha=0.0)[source]
```

Ensemble Learning with Scikit-Learn

- Boosting:

- https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.AdaBoostClassifier.html

sklearn.ensemble.AdaBoostClassifier

```
class sklearn.ensemble.AdaBoostClassifier(estimator=None, *, n_estimators=50, learning_rate=1.0,
algorithm='SAMME.R', random_state=None)                                                    [source]
```

- https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.AdaBoostRegressor.html

sklearn.ensemble.AdaBoostRegressor

```
class sklearn.ensemble.AdaBoostRegressor(estimator=None, *, n_estimators=50, learning_rate=1.0, loss='linear',
random_state=None)                                                                         [source]
```

Ensemble Learning with Scikit-Learn

- Stacking:

- https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.StackingClassifier.html

sklearn.ensemble.StackingClassifier

```
class sklearn.ensemble.StackingClassifier(estimators, final_estimator=None, *, cv=None, stack_method='auto',
n_jobs=None, passthrough=False, verbose=0)                                                      [source]
```

- https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.StackingRegressor.html

sklearn.ensemble.StackingRegressor

```
class sklearn.ensemble.StackingRegressor(estimators, final_estimator=None, *, cv=None, n_jobs=None,
passthrough=False, verbose=0)                                                                   [source]
```

Ensemble Learning with Scikit-Learn

- Voting:

- https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingClassifier.html

sklearn.ensemble.VotingClassifier

```
class sklearn.ensemble.VotingClassifier(estimators, *, voting='hard', weights=None, n_jobs=None,
flatten_transform=True, verbose=False)                                                    [source]
```

- https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingRegressor.html

sklearn.ensemble.VotingRegressor

```
class sklearn.ensemble.VotingRegressor(estimators, *, weights=None, n_jobs=None, verbose=False)    [source]
```

Hyperparameter Optimization with Scikit-Learn

- ## Grid Search:

  - https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html

  sklearn.model_selection.GridSearchCV

  ```
  class sklearn.model_selection.GridSearchCV(estimator, param_grid, *, scoring=None, n_jobs=None, refit=True,
  cv=None, verbose=0, pre_dispatch='2*n_jobs', error_score=nan, return_train_score=False)
  ```
  [source]

- ## Randomized Search:

  - https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.RandomizedSearchCV.html

  sklearn.model_selection.RandomizedSearchCV

  ```
  class sklearn.model_selection.RandomizedSearchCV(estimator, param_distributions, *, n_iter=10, scoring=None,
  n_jobs=None, refit=True, cv=None, verbose=0, pre_dispatch='2*n_jobs', random_state=None, error_score=nan,
  return_train_score=False)
  ```
  [source]

Exercises:

- Notebooks on the github repository:
  - Notebook with exercises:
    - exercises/session18/exercises.ipynb