산업인공지능개론

XGBoost, LightGBM

2023254006
이선경

CONTENS

01. XGBoost

코드 구현

```
import numpy as np
import pandas as pd
from sklearn.datasets import load_boston
from sklearn.metrics import mean_squared_error
from sklearn.model_selection import train_test_split
import xgboost as xgb
boston = load_boston()
data = pd.DataFrame(boston.data)
data.columns = boston.feature_names
data['PRICE'] = boston.target
print(data.head())
X, y = data.iloc[:,:-1],data.iloc[:,-1]
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=123)
xg_reg = xgb.XGBRegressor(objective ='reg:squarederror', colsample_bytree = 0.3,
                          learning_rate = 0.1, max_depth = 5, alpha = 10, n_estimators = 10)
xg_reg.fit(X_train, y_train)
preds = xg_reg.predict(X_test)
rmse = np.sqrt(mean_squared_error(y_test, preds))
print("RMSE: %f" % (rmse))
```

01. XGBoost

결과

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT	PRICE
0	0.00632	18.0	2.31	0.0	0.538	6.575	65.2	4.0900	1.0	296.0	15.3	396.90	4.98	24.0
1	0.02731	0.0	7.07	0.0	0.469	6.421	78.9	4.9671	2.0	242.0	17.8	396.90	9.14	21.6
2	0.02729	0.0	7.07	0.0	0.469	7.185	61.1	4.9671	2.0	242.0	17.8	392.83	4.03	34.7
3	0.03237	0.0	2.18	0.0	0.458	6.998	45.8	6.0622	3.0	222.0	18.7	394.63	2.94	33.4
4	0.06905	0.0	2.18	0.0	0.458	7.147	54.2	6.0622	3.0	222.0	18.7	396.90	5.33	36.2

RMSE: 10.423243

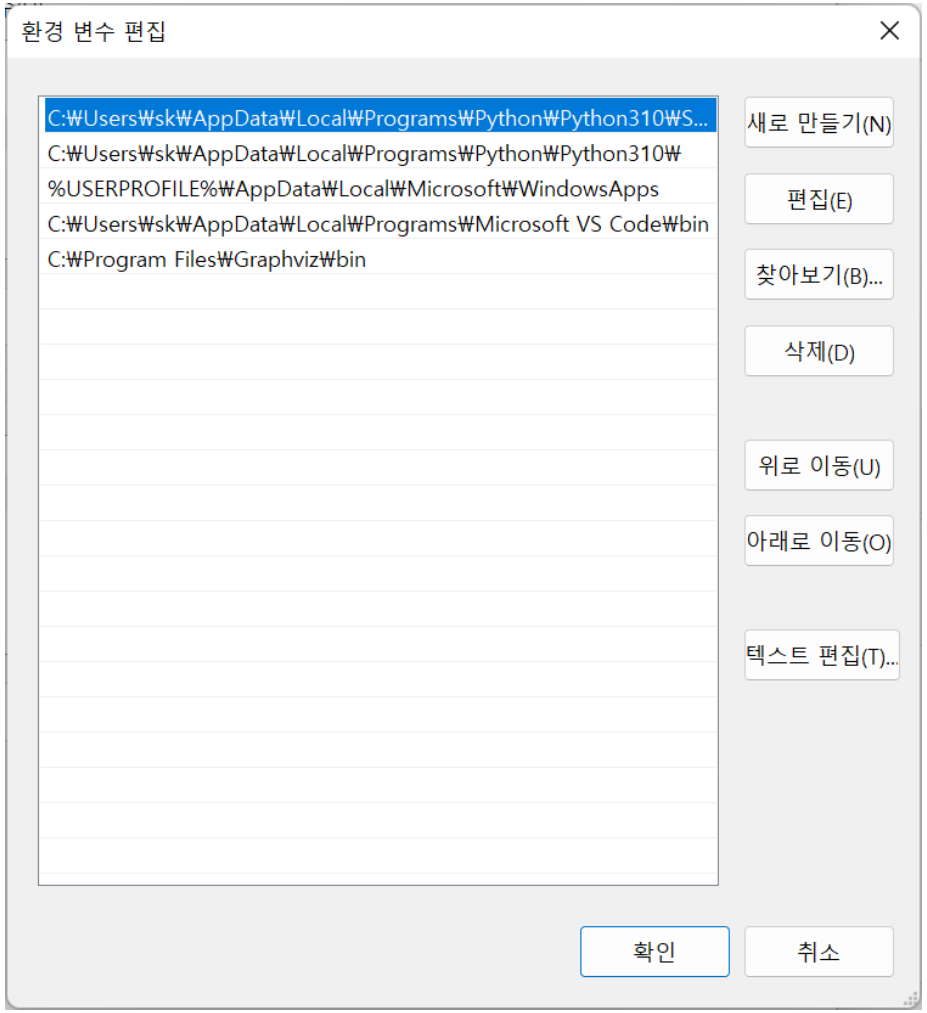
02. LightGBM

코드 구현

```python
import matplotlib.pyplot as plt
from lightgbm import LGBMClassifier, LGBMRegressor
from lightgbm import plot_importance, plot_metric, plot_tree
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
from sklearn.model_selection import cross_validate

iris = load_iris()
X_train, X_test, y_train, y_test = train_test_split(iris.data, iris.target, test_size=0.2, random_state=123)
lgbmc = LGBMClassifier(n_estimators=400)
evals = [(X_test, y_test)]
lgbmc.fit(X_train, y_train, early_stopping_rounds=100, eval_metric='logloss', eval_set=evals, verbose=True)
preds = lgbmc.predict(X_test)

cross_val = cross_validate(
    estimator=lgbmc,
    X=iris.data, y=iris.target,
    cv=5
)

print('avg fit time: {} (+/- {})'.format(cross_val['fit_time']. mean(), cross_val['fit_time'].std()))
print('avg score time: {} (+/- {})'.format(cross_val['score_time']. mean(), cross_val['score_time'].std()))
print('avg test score: {} (+/- {})'.format(cross_val['test_score'].mean(), cross_val['test_score'].std()))

fig, ax = plt.subplots(figsize=(12, 8))
plot_metric(lgbmc)
plot_importance(lgbmc, figsize=(10,12))
plot_tree(lgbmc, ax=ax, figsize=(28,24))
plt.show()
```

02. LightGBM

Graphviz 설치 후 환경 변수 Graphviz 경로 추가



02. LightGBM

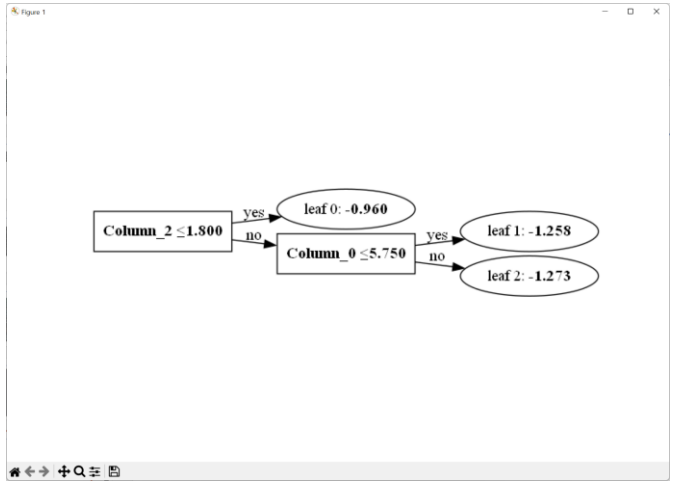
결과

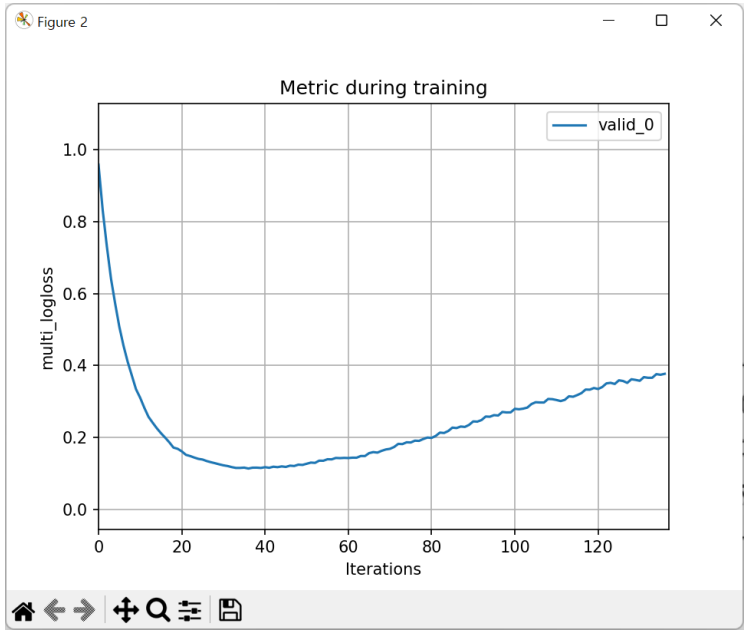
```
[1]     valid_0's multi_logloss: 0.95847
[2]     valid_0's multi_logloss: 0.832184
[3]     valid_0's multi_logloss: 0.731164
[4]     valid_0's multi_logloss: 0.641056
[5]     valid_0's multi_logloss: 0.571726
[6]     valid_0's multi_logloss: 0.507286
[7]     valid_0's multi_logloss: 0.454933
[8]     valid_0's multi_logloss: 0.410205
[9]     valid_0's multi_logloss: 0.372194
[10]    valid_0's multi_logloss: 0.333919
[11]    valid_0's multi_logloss: 0.310212
[12]    valid_0's multi_logloss: 0.282326
[13]    valid_0's multi_logloss: 0.257165
[14]    valid_0's multi_logloss: 0.240836
[15]    valid_0's multi_logloss: 0.225383
[16]    valid_0's multi_logloss: 0.211583
[17]    valid_0's multi_logloss: 0.199289
[18]    valid_0's multi_logloss: 0.186269
[19]    valid_0's multi_logloss: 0.171556
[20]    valid_0's multi_logloss: 0.168245
[21]    valid_0's multi_logloss: 0.161065
[22]    valid_0's multi_logloss: 0.151371
[23]    valid_0's multi_logloss: 0.148081
[24]    valid_0's multi_logloss: 0.143843
[25]    valid_0's multi_logloss: 0.140169
[26]    valid_0's multi_logloss: 0.138303
[27]    valid_0's multi_logloss: 0.134058
[28]    valid_0's multi_logloss: 0.130884
[29]    valid_0's multi_logloss: 0.128082
[30]    valid_0's multi_logloss: 0.124975
[31]    valid_0's multi_logloss: 0.122225
[32]    valid_0's multi_logloss: 0.120298
[33]    valid_0's multi_logloss: 0.117257
[34]    valid_0's multi_logloss: 0.115021
[35]    valid_0's multi_logloss: 0.115037
[36]    valid_0's multi_logloss: 0.115831
[37]    valid_0's multi_logloss: 0.113318
[38]    valid_0's multi_logloss: 0.115651
[39]    valid_0's multi_logloss: 0.115772
[40]    valid_0's multi_logloss: 0.114953
[41]    valid_0's multi_logloss: 0.117056
[42]    valid_0's multi_logloss: 0.115412
[43]    valid_0's multi_logloss: 0.118359
[44]    valid_0's multi_logloss: 0.117129
[45]    valid_0's multi_logloss: 0.119174
[46]    valid_0's multi_logloss: 0.117789
[47]    valid_0's multi_logloss: 0.121333
[48]    valid_0's multi_logloss: 0.120375
[49]    valid_0's multi_logloss: 0.124128
[50]    valid_0's multi_logloss: 0.123394
[109]   valid_0's multi_logloss: 0.30682
[110]   valid_0's multi_logloss: 0.306206
```

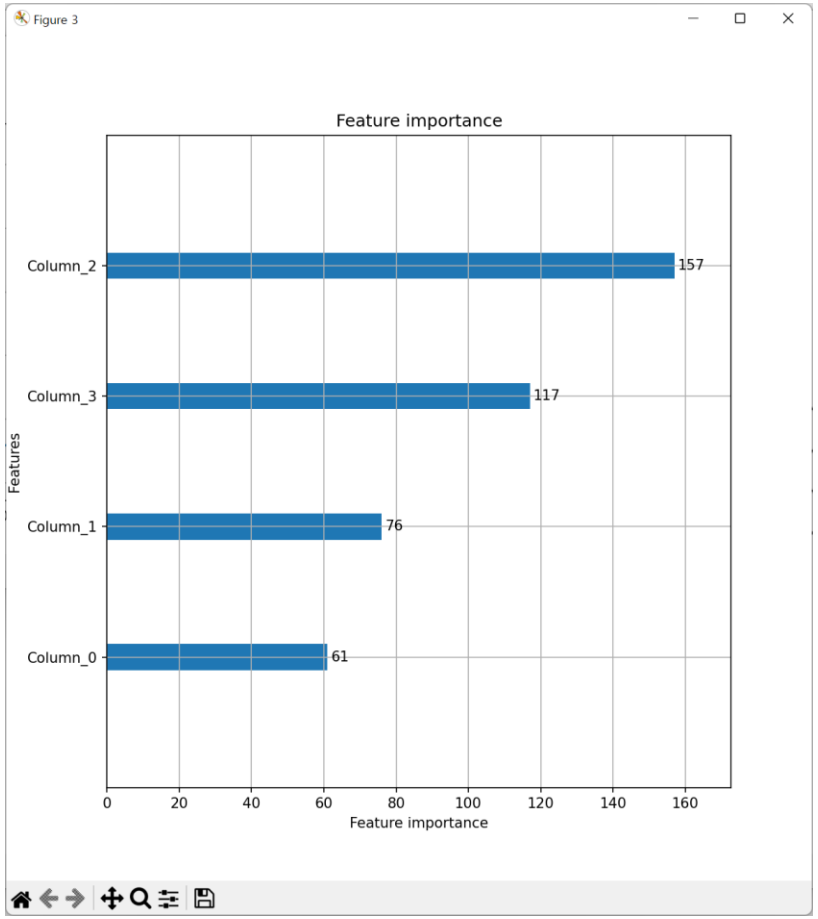
```
[111]   valid_0's multi_logloss: 0.303895
[112]   valid_0's multi_logloss: 0.300907
[113]   valid_0's multi_logloss: 0.304274
[114]   valid_0's multi_logloss: 0.314218
[115]   valid_0's multi_logloss: 0.312988
[116]   valid_0's multi_logloss: 0.317589
[117]   valid_0's multi_logloss: 0.323073
[118]   valid_0's multi_logloss: 0.333026
[119]   valid_0's multi_logloss: 0.332652
[120]   valid_0's multi_logloss: 0.337212
[121]   valid_0's multi_logloss: 0.334481
[122]   valid_0's multi_logloss: 0.340022
[123]   valid_0's multi_logloss: 0.350061
[124]   valid_0's multi_logloss: 0.351676
[125]   valid_0's multi_logloss: 0.348515
[126]   valid_0's multi_logloss: 0.358595
[127]   valid_0's multi_logloss: 0.356737
[128]   valid_0's multi_logloss: 0.351512
[129]   valid_0's multi_logloss: 0.361591
[130]   valid_0's multi_logloss: 0.35978
[131]   valid_0's multi_logloss: 0.357317
[132]   valid_0's multi_logloss: 0.367439
[133]   valid_0's multi_logloss: 0.365665
[134]   valid_0's multi_logloss: 0.365745
[135]   valid_0's multi_logloss: 0.375832
[136]   valid_0's multi_logloss: 0.374115
[137]   valid_0's multi_logloss: 0.376748
avg fit time: 0.07919144630432129 (+/- 0.014747029726158264)
avg score time: 0.001833200454711914 (+/- 0.0006676939103703261)
avg test score: 0.9600000000000002 (+/- 0.04898979485566355)
```

02. LightGBM

결과







02. LightGBM

Python 설치 패키지