

위스콘신 유방암 데이터 셋 - LightGBM 모델 구현

학습 목표

- LightGBM 모델을 이용하여 학습과 예측을 수행해 본다.
- 예측을 수행한 결과에 대해 평가를 수행해 본다.(교차검증)

학습 내용

- LightGBM 모델 만들기 - 위스콘신 유방암 데이터 셋
- 평가해보기

목차

01 데이터 로드 및 전처리

02 LightGBM 모델

03 피쳐 중요도 시각화

04 교차 검증

05 모델 평가 - 교차 검증, GridSearchCV()

01 데이터 로드 및 전처리

목차로 이동하기

- 라이브러리가 없다고 확인되면 설치를 진행
 - (예) pip install lightgbm

```
In [1]: import pandas as pd
from sklearn.datasets import load_breast_cancer
import matplotlib.pyplot as plt
import matplotlib
```

```
import lightgbm as lgbm
import sklearn as sk
from lightgbm import LGBMClassifier
from sklearn.model_selection import train_test_split
```

```
In [2]: print("LightGBM : ", lgbm.__version__)
print("Scikit-learn : ", sk.__version__)
print("pandas : ", pd.__version__)
```

```
LightGBM : 3.3.5
Scikit-learn : 1.0.2
pandas : 1.4.4
```

```
In [3]: cancer = load_breast_cancer()

cancer_df = pd.DataFrame(cancer.data, columns=cancer.feature_names)
cancer_df.head()
```

Out[3]:

	mean radius	mean texture	mean perimeter	mean area	mean smoothness	mean compactness	mean concavity	mean concave points	mean symmetry
0	17.99	10.38	122.80	1001.0	0.11840	0.27760	0.3001	0.14710	0.2419
1	20.57	17.77	132.90	1326.0	0.08474	0.07864	0.0869	0.07017	0.1812
2	19.69	21.25	130.00	1203.0	0.10960	0.15990	0.1974	0.12790	0.2069
3	11.42	20.38	77.58	386.1	0.14250	0.28390	0.2414	0.10520	0.2597
4	20.29	14.34	135.10	1297.0	0.10030	0.13280	0.1980	0.10430	0.1809

5 rows × 30 columns

```
In [4]: print( cancer_df.shape )  
(569, 30)
```

데이터 설명

- 위스콘신 유방암 데이터 세트는 유방암의 악성 종양, 양성 종양 여부를 결정하는 이진 분류
- 종양의 크기, 모양 등의 형태와 관련한 많은 피처를 가지고 있음.
- 569개의 행과, 30개의 피처로 이루어진 데이터
- null 값이 없음. 값들은 실수로 되어 있음.

데이터 나누기

```
In [21]: # 피처와 레이블을 지정.  
X = cancer_df[:,  
y = cancer.target  
  
X.shape, y.shape
```

Out[21]: ((569, 30), (569,))

```
In [22]: X_train , X_test, y_train, y_test = train_test_split(X, y,  
test_size=0.2, random_state=0)  
  
X_train.shape, X_test.shape, y_train.shape, y_test.shape
```

Out[22]: ((455, 30), (114, 30), (455,), (114,))

02 LightGBM 모델

목차로 이동하기

- LightGBM은 XGBoost와 부스팅 계열 알고리즘에서 가장 각광을 받고 있음.
- LightGBM은 가장 큰 장점은 XGBoost보다 학습에 걸리는 시간이 훨씬 적다. 메모리 사용량도 적다.
- LightGBM과 XGBoost의 예측 성능은 별다른 차이가 없음.
- LightGBM이 XGBoost보다 2년 후에 만들어짐.

- [단점] 적은 데이터 셋일 경우, 과적합이 발생할 가능성이 있음. (문서상에는 10000건 이하의 데이터로 기술하고 있음.)
- lightgbm 설치
 - pip install lightgbm

LightGBM의 파이썬 패키지인 lightgbm에서 LGBMClassifier 불러오기

```
In [23]: from lightgbm import LGBMClassifier
```

```
In [24]: # 모델 선택
model_lgbm = LGBMClassifier(n_estimators= 100)

# Lgbm은 중간에 조기 중단이 가능.
evals = [(X_test, y_test)]

# 모델 학습
model_lgbm.fit(X_train, y_train, early_stopping_rounds=100,
               eval_metric='logloss',
               eval_set=evals,
               verbose=True)

# 모델을 활용한 예측
preds = model_lgbm.predict(X_test)
pred_proba = model_lgbm.predict_proba(X_test)[: , 1]
```

[1]	valid_0's binary_logloss: 0.602575
[2]	valid_0's binary_logloss: 0.536123
[3]	valid_0's binary_logloss: 0.482224
[4]	valid_0's binary_logloss: 0.439121
[5]	valid_0's binary_logloss: 0.399484
[6]	valid_0's binary_logloss: 0.360741
[7]	valid_0's binary_logloss: 0.329635
[8]	valid_0's binary_logloss: 0.300259
[9]	valid_0's binary_logloss: 0.279599
[10]	valid_0's binary_logloss: 0.259173
[11]	valid_0's binary_logloss: 0.239127
[12]	valid_0's binary_logloss: 0.222309
[13]	valid_0's binary_logloss: 0.209351
[14]	valid_0's binary_logloss: 0.194316
[15]	valid_0's binary_logloss: 0.185282
[16]	valid_0's binary_logloss: 0.173675
[17]	valid_0's binary_logloss: 0.165308
[18]	valid_0's binary_logloss: 0.155769
[19]	valid_0's binary_logloss: 0.14865
[20]	valid_0's binary_logloss: 0.140995
[21]	valid_0's binary_logloss: 0.134334
[22]	valid_0's binary_logloss: 0.126574
[23]	valid_0's binary_logloss: 0.119264
[24]	valid_0's binary_logloss: 0.114799
[25]	valid_0's binary_logloss: 0.110568
[26]	valid_0's binary_logloss: 0.106157
[27]	valid_0's binary_logloss: 0.100816
[28]	valid_0's binary_logloss: 0.0978631
[29]	valid_0's binary_logloss: 0.0945873
[30]	valid_0's binary_logloss: 0.0916942
[31]	valid_0's binary_logloss: 0.0863612
[32]	valid_0's binary_logloss: 0.0822965
[33]	valid_0's binary_logloss: 0.078824
[34]	valid_0's binary_logloss: 0.0752847
[35]	valid_0's binary_logloss: 0.0718233
[36]	valid_0's binary_logloss: 0.0696371
[37]	valid_0's binary_logloss: 0.0676798
[38]	valid_0's binary_logloss: 0.0665177
[39]	valid_0's binary_logloss: 0.0660656
[40]	valid_0's binary_logloss: 0.064516
[41]	valid_0's binary_logloss: 0.0634248
[42]	valid_0's binary_logloss: 0.0616478
[43]	valid_0's binary_logloss: 0.0603263
[44]	valid_0's binary_logloss: 0.0598075
[45]	valid_0's binary_logloss: 0.058935
[46]	valid_0's binary_logloss: 0.0594708
[47]	valid_0's binary_logloss: 0.0575433
[48]	valid_0's binary_logloss: 0.05764
[49]	valid_0's binary_logloss: 0.058022
[50]	valid_0's binary_logloss: 0.056115
[51]	valid_0's binary_logloss: 0.0551967
[52]	valid_0's binary_logloss: 0.0546137
[53]	valid_0's binary_logloss: 0.055392
[54]	valid_0's binary_logloss: 0.0550331
[55]	valid_0's binary_logloss: 0.0550406
[56]	valid_0's binary_logloss: 0.0551554
[57]	valid_0's binary_logloss: 0.0550553
[58]	valid_0's binary_logloss: 0.0549359
[59]	valid_0's binary_logloss: 0.056094
[60]	valid_0's binary_logloss: 0.0571328
[61]	valid_0's binary_logloss: 0.0580245
[62]	valid_0's binary_logloss: 0.0574484
[63]	valid_0's binary_logloss: 0.0577891
[64]	valid_0's binary_logloss: 0.0592917
[65]	valid_0's binary_logloss: 0.0597466

```
[66] valid_0's binary_logloss: 0.0597562
[67] valid_0's binary_logloss: 0.0608003
[68] valid_0's binary_logloss: 0.0599771
[69] valid_0's binary_logloss: 0.060662
[70] valid_0's binary_logloss: 0.0610663
[71] valid_0's binary_logloss: 0.0617392
[72] valid_0's binary_logloss: 0.0611657
[73] valid_0's binary_logloss: 0.0607031
[74] valid_0's binary_logloss: 0.0607682
[75] valid_0's binary_logloss: 0.0618635
[76] valid_0's binary_logloss: 0.0632427
[77] valid_0's binary_logloss: 0.0641161
[78] valid_0's binary_logloss: 0.0648781
[79] valid_0's binary_logloss: 0.064908
[80] valid_0's binary_logloss: 0.0650122
[81] valid_0's binary_logloss: 0.0651241
[82] valid_0's binary_logloss: 0.0651113
[83] valid_0's binary_logloss: 0.0650562
[84] valid_0's binary_logloss: 0.0642698
[85] valid_0's binary_logloss: 0.0629834
[86] valid_0's binary_logloss: 0.0632386
[87] valid_0's binary_logloss: 0.0623831
[88] valid_0's binary_logloss: 0.0611975
[89] valid_0's binary_logloss: 0.060535
[90] valid_0's binary_logloss: 0.0604596
[91] valid_0's binary_logloss: 0.0594299
[92] valid_0's binary_logloss: 0.0594986
[93] valid_0's binary_logloss: 0.0599838
[94] valid_0's binary_logloss: 0.05974
[95] valid_0's binary_logloss: 0.0606709
[96] valid_0's binary_logloss: 0.0607891
[97] valid_0's binary_logloss: 0.062165
[98] valid_0's binary_logloss: 0.0608996
[99] valid_0's binary_logloss: 0.062041
[100] valid_0's binary_logloss: 0.0608407
```

```
C:\Users\daniel_wj\Anaconda3\lib\site-packages\lightgbm\sklearn.py:726: UserWarning:
'early_stopping_rounds' argument is deprecated and will be removed in a future release of LightGBM. Pass 'early_stopping()' callback via 'callbacks' argument instead.
_log_warning("'early_stopping_rounds' argument is deprecated and will be removed in a future release of LightGBM. ")
C:\Users\daniel_wj\Anaconda3\lib\site-packages\lightgbm\sklearn.py:736: UserWarning:
'verbose' argument is deprecated and will be removed in a future release of LightGBM. Pass 'log_evaluation()' callback via 'callbacks' argument instead.
_log_warning("'verbose' argument is deprecated and will be removed in a future release of LightGBM. ")
```

- 조기 중단으로 계속 수행하지 않고, 중간에 중단함.

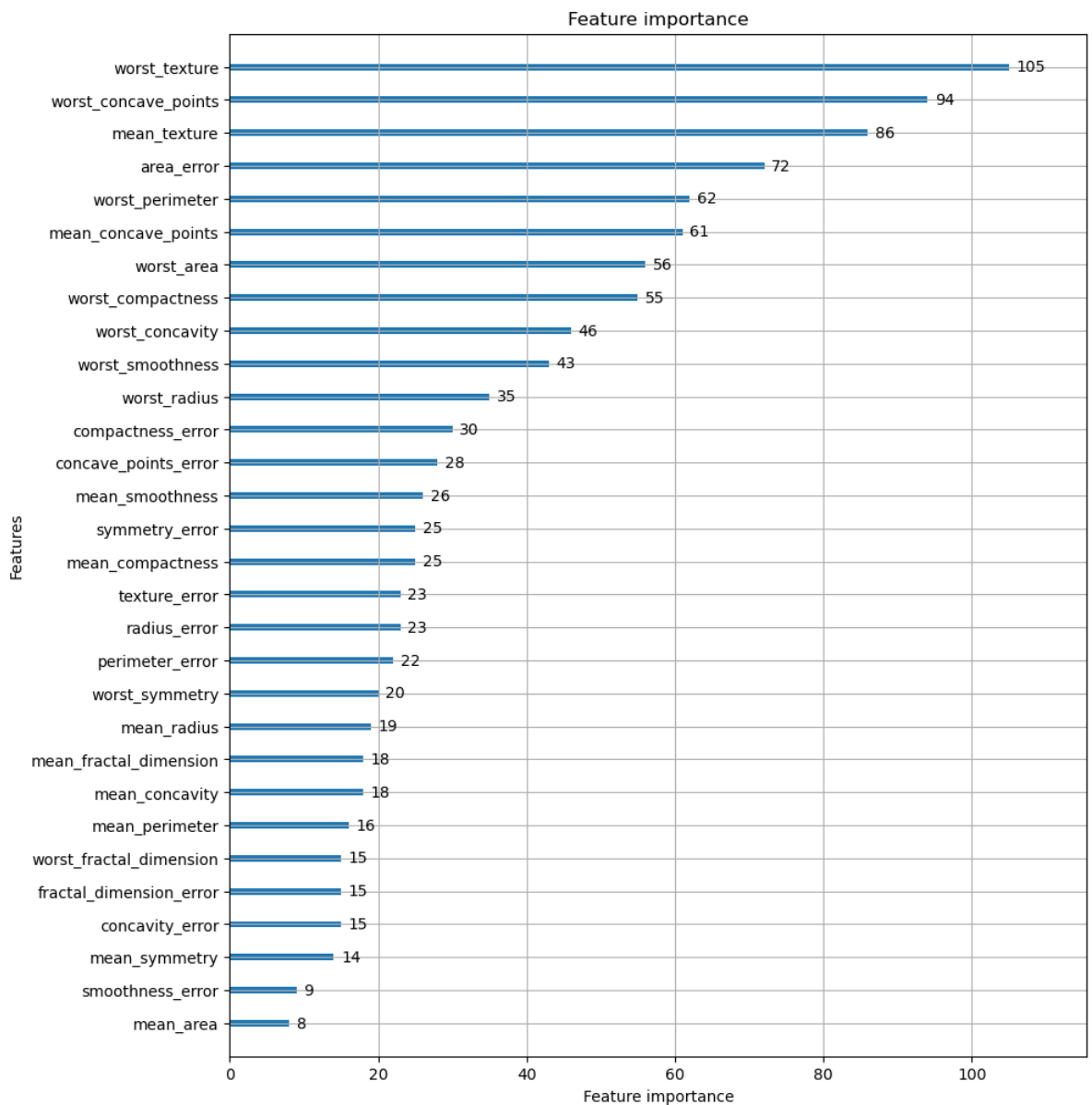
03 피쳐 중요도 시각화

목차로 이동하기

```
In [25]: from lightgbm import plot_importance
import matplotlib.pyplot as plt
```

```
fig, ax = plt.subplots(figsize=(10,12))
plot_importance(model_lgbm, ax=ax)
```

```
Out[25]: <AxesSubplot:title={'center':'Feature importance'}, xlabel='Feature importance', ylabel='Features'>
```



04 모델 평가 - 교차 검증

목차로 이동하기

```
In [15]: from sklearn.model_selection import cross_val_score
```

```
In [18]: def model_val_lgbm(model_name):
    model = model_name

    # Lgbm은 중간에 조기 중단이 가능.
    evals = [(X_test, y_test)]

    # 모델 학습
    model.fit(X_train, y_train, early_stopping_rounds=100,
              eval_metric='logloss',
              eval_set=evals,
              verbose=False) # True는 학습 과정이 보임

    score = cross_val_score(model, X_test, y_test, cv=5, scoring="roc_auc")

    print("auc :", score)
    m_score = score.mean() # 절대값
```

```
print("auc 평균(cv=5) : ", m_score)
return m_score
```

```
In [19]: # 모델 선택
model_lgbm = LGBMClassifier(n_estimators= 100,
                             learning_rate=0.1,
                             max_depth = 3)

score = model_val_lgbm(model_lgbm)
score

auc : [0.95384615 0.99230769 1.          0.98412698 0.97435897]
auc 평균(cv=5) : 0.980927960927961
```

C:\Users\Wdaniel_wj\Anaconda3\lib\site-packages\lightgbm\sklearn.py:726: UserWarning: 'early_stopping_rounds' argument is deprecated and will be removed in a future release of LightGBM. Pass 'early_stopping()' callback via 'callbacks' argument instead.
 _log_warning("'early_stopping_rounds' argument is deprecated and will be removed in a future release of LightGBM. ")
C:\Users\Wdaniel_wj\Anaconda3\lib\site-packages\lightgbm\sklearn.py:736: UserWarning: 'verbose' argument is deprecated and will be removed in a future release of LightGBM. Pass 'log_evaluation()' callback via 'callbacks' argument instead.
 _log_warning("'verbose' argument is deprecated and will be removed in a future release of LightGBM. ")

```
Out[19]: 0.980927960927961
```

05 모델 평가 - 교차 검증, GridSearchCV()

목차로 이동하기

```
In [35]: param_grid = {'max_depth':[3,5,7],
                       'learning_rate':[0.1, 0.01, 0.001],
                       'n_estimators':[100, 300, 500]}
print("매개 변수 그리드 : \n", param_grid)

매개 변수 그리드 :
{'max_depth': [3, 5, 7], 'learning_rate': [0.1, 0.01, 0.001], 'n_estimators': [100, 300, 500]}
```

```
In [36]: from sklearn.model_selection import GridSearchCV
```

```
In [37]: grid_search = GridSearchCV(LGBMClassifier(),
                                   param_grid, cv=5, return_train_score=True, scoring="roc_
grid_search
```

```
Out[37]: GridSearchCV(cv=5, estimator=LGBMClassifier(),
                      param_grid={'learning_rate': [0.1, 0.01, 0.001],
                                   'max_depth': [3, 5, 7],
                                   'n_estimators': [100, 300, 500]},
                      return_train_score=True, scoring='roc_auc')
```

```
In [38]: # 찾기 학습
evals = [(X_test, y_test)]
grid_search.fit(X_train, y_train, early_stopping_rounds=100,
                eval_metric='logloss',
                eval_set=evals,
                verbose=False) # True는 학습 과정이 보임)
```

[illegible]


```
C:\Users\daniel_wj\Anaconda3\Lib\site-packages\lightgbm\sklearn.py:736: UserWarning:  
'verbose' argument is deprecated and will be removed in a future release of LightGBM.  
Pass 'log_evaluation()' callback via 'callbacks' argument instead.
```

```
_log_warning("'verbose' argument is deprecated and will be removed in a future rel  
ease of LightGBM. ")  
C:\Users\daniel_wj\Anaconda3\Lib\site-packages\lightgbm\sklearn.py:726: UserWarning:  
'early_stopping_rounds' argument is deprecated and will be removed in a future relea  
se of LightGBM. Pass 'early_stopping()' callback via 'callbacks' argument instead.
```

```
_log_warning("'early_stopping_rounds' argument is deprecated and will be removed i  
n a future release of LightGBM. ")  
C:\Users\daniel_wj\Anaconda3\Lib\site-packages\lightgbm\sklearn.py:736: UserWarning:  
'verbose' argument is deprecated and will be removed in a future release of LightGB  
M. Pass 'log_evaluation()' callback via 'callbacks' argument instead.
```

```
_log_warning("'verbose' argument is deprecated and will be removed in a future rel  
ease of LightGBM. ")  
C:\Users\daniel_wj\Anaconda3\Lib\site-packages\lightgbm\sklearn.py:726: UserWarning:  
'early_stopping_rounds' argument is deprecated and will be removed in a future relea  
se of LightGBM. Pass 'early_stopping()' callback via 'callbacks' argument instead.
```

```
_log_warning("'early_stopping_rounds' argument is deprecated and will be removed i  
n a future release of LightGBM. ")  
C:\Users\daniel_wj\Anaconda3\Lib\site-packages\lightgbm\sklearn.py:736: UserWarning:  
'verbose' argument is deprecated and will be removed in a future release of LightGB  
M. Pass 'log_evaluation()' callback via 'callbacks' argument instead.
```

```
_log_warning("'verbose' argument is deprecated and will be removed in a future rel  
ease of LightGBM. ")  
C:\Users\daniel_wj\Anaconda3\Lib\site-packages\lightgbm\sklearn.py:726: UserWarning:  
'early_stopping_rounds' argument is deprecated and will be removed in a future relea  
se of LightGBM. Pass 'early_stopping()' callback via 'callbacks' argument instead.
```

```
_log_warning("'early_stopping_rounds' argument is deprecated and will be removed i  
n a future release of LightGBM. ")  
C:\Users\daniel_wj\Anaconda3\Lib\site-packages\lightgbm\sklearn.py:736: UserWarning:  
'verbose' argument is deprecated and will be removed in a future release of LightGB  
M. Pass 'log_evaluation()' callback via 'callbacks' argument instead.
```

```
_log_warning("'verbose' argument is deprecated and will be removed in a future rel  
ease of LightGBM. ")  
C:\Users\daniel_wj\Anaconda3\Lib\site-packages\lightgbm\sklearn.py:726: UserWarning:  
'early_stopping_rounds' argument is deprecated and will be removed in a future relea  
se of LightGBM. Pass 'early_stopping()' callback via 'callbacks' argument instead.
```

```
_log_warning("'early_stopping_rounds' argument is deprecated and will be removed i  
n a future release of LightGBM. ")  
C:\Users\daniel_wj\Anaconda3\Lib\site-packages\lightgbm\sklearn.py:736: UserWarning:  
'verbose' argument is deprecated and will be removed in a future release of LightGB  
M. Pass 'log_evaluation()' callback via 'callbacks' argument instead.
```

```
_log_warning("'verbose' argument is deprecated and will be removed in a future rel  
ease of LightGBM. ")
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```

C:\Users\Wdaniel_wj\Wanaconda3\lib\site-packages\lightgbm\sklearn.py:726: UserWarning:
'early_stopping_rounds' argument is deprecated and will be removed in a future release of LightGBM. Pass 'early_stopping()' callback via 'callbacks' argument instead.
  _log_warning("'early_stopping_rounds' argument is deprecated and will be removed in a future release of LightGBM. ")
C:\Users\Wdaniel_wj\Wanaconda3\lib\site-packages\lightgbm\sklearn.py:736: UserWarning:
'verbos' argument is deprecated and will be removed in a future release of LightGBM. Pass 'log_evaluation()' callback via 'callbacks' argument instead.
  _log_warning("'verbos' argument is deprecated and will be removed in a future release of LightGBM. ")
C:\Users\Wdaniel_wj\Wanaconda3\lib\site-packages\lightgbm\sklearn.py:726: UserWarning:
'early_stopping_rounds' argument is deprecated and will be removed in a future release of LightGBM. Pass 'early_stopping()' callback via 'callbacks' argument instead.
  _log_warning("'early_stopping_rounds' argument is deprecated and will be removed in a future release of LightGBM. ")
C:\Users\Wdaniel_wj\Wanaconda3\lib\site-packages\lightgbm\sklearn.py:736: UserWarning:
'verbos' argument is deprecated and will be removed in a future release of LightGBM. Pass 'log_evaluation()' callback via 'callbacks' argument instead.
  _log_warning("'verbos' argument is deprecated and will be removed in a future release of LightGBM. ")
C:\Users\Wdaniel_wj\Wanaconda3\lib\site-packages\lightgbm\sklearn.py:726: UserWarning:
'early_stopping_rounds' argument is deprecated and will be removed in a future release of LightGBM. Pass 'early_stopping()' callback via 'callbacks' argument instead.
  _log_warning("'early_stopping_rounds' argument is deprecated and will be removed in a future release of LightGBM. ")
C:\Users\Wdaniel_wj\Wanaconda3\lib\site-packages\lightgbm\sklearn.py:736: UserWarning:
'verbos' argument is deprecated and will be removed in a future release of LightGBM. Pass 'log_evaluation()' callback via 'callbacks' argument instead.
  _log_warning("'verbos' argument is deprecated and will be removed in a future release of LightGBM. ")
C:\Users\Wdaniel_wj\Wanaconda3\lib\site-packages\lightgbm\sklearn.py:726: UserWarning:
'early_stopping_rounds' argument is deprecated and will be removed in a future release of LightGBM. Pass 'early_stopping()' callback via 'callbacks' argument instead.
  _log_warning("'early_stopping_rounds' argument is deprecated and will be removed in a future release of LightGBM. ")
C:\Users\Wdaniel_wj\Wanaconda3\lib\site-packages\lightgbm\sklearn.py:736: UserWarning:
'verbos' argument is deprecated and will be removed in a future release of LightGBM. Pass 'log_evaluation()' callback via 'callbacks' argument instead.
  _log_warning("'verbos' argument is deprecated and will be removed in a future release of LightGBM. ")
C:\Users\Wdaniel_wj\Wanaconda3\lib\site-packages\lightgbm\sklearn.py:726: UserWarning:
'early_stopping_rounds' argument is deprecated and will be removed in a future release of LightGBM. Pass 'early_stopping()' callback via 'callbacks' argument instead.
  _log_warning("'early_stopping_rounds' argument is deprecated and will be removed in a future release of LightGBM. ")
C:\Users\Wdaniel_wj\Wanaconda3\lib\site-packages\lightgbm\sklearn.py:736: UserWarning:
'verbos' argument is deprecated and will be removed in a future release of LightGBM. Pass 'log_evaluation()' callback via 'callbacks' argument instead.
  _log_warning("'verbos' argument is deprecated and will be removed in a future release of LightGBM. ")
C:\Users\Wdaniel_wj\Wanaconda3\lib\site-packages\lightgbm\sklearn.py:726: UserWarning:
'early_stopping_rounds' argument is deprecated and will be removed in a future release of LightGBM. Pass 'early_stopping()' callback via 'callbacks' argument instead.
  _log_warning("'early_stopping_rounds' argument is deprecated and will be removed in a future release of LightGBM. ")
C:\Users\Wdaniel_wj\Wanaconda3\lib\site-packages\lightgbm\sklearn.py:736: UserWarning:
'verbos' argument is deprecated and will be removed in a future release of LightGBM. Pass 'log_evaluation()' callback via 'callbacks' argument instead.
  _log_warning("'verbos' argument is deprecated and will be removed in a future release of LightGBM. ")

```

```

Out[38]: GridSearchCV(cv=5, estimator=LGBMClassifier(),
                    param_grid={'learning_rate': [0.1, 0.01, 0.001],
                                'max_depth': [3, 5, 7],
                                'n_estimators': [100, 300, 500]},
                    return_train_score=True, scoring='roc_auc')

```

```
In [39]: best_model = grid_search.best_estimator_  
best_params = grid_search.best_params_
```

```
best_model, best_params
```

```
Out[39]: (LGBMClassifier(max_depth=5, n_estimators=300),  
{'learning_rate': 0.1, 'max_depth': 5, 'n_estimators': 300})
```

```
In [40]: print("최적 매개변수 :", grid_search.best_params_)  
print("최고 교차 검증 점수 : {:.4f}".format(grid_search.best_score_))
```

```
최적 매개변수 : {'learning_rate': 0.1, 'max_depth': 5, 'n_estimators': 300}  
최고 교차 검증 점수 : 0.9935
```

정리

- gridSearchCV()를 이용하여 다양한 파라미터의 성능을 확인해 볼 수 있습니다.