AI

XG Boost VS Light GBM

- 91714167 유재겸

Contents

01. 사전지식

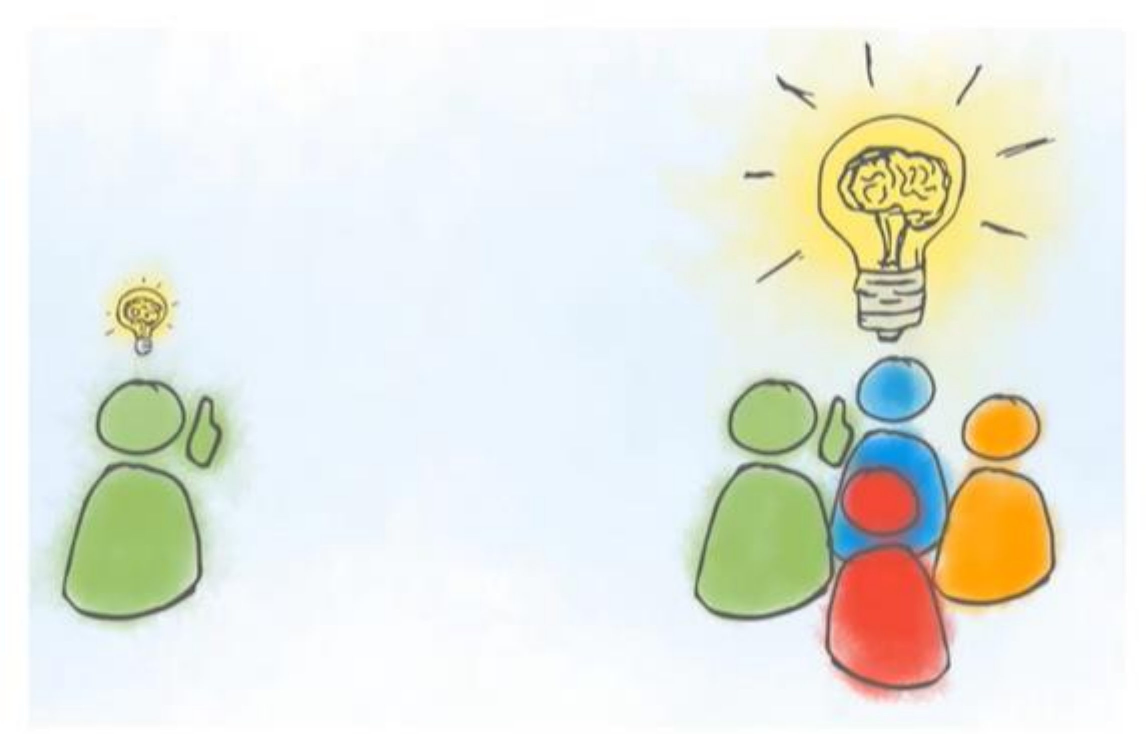
02. XG Boost

03. Light GBM

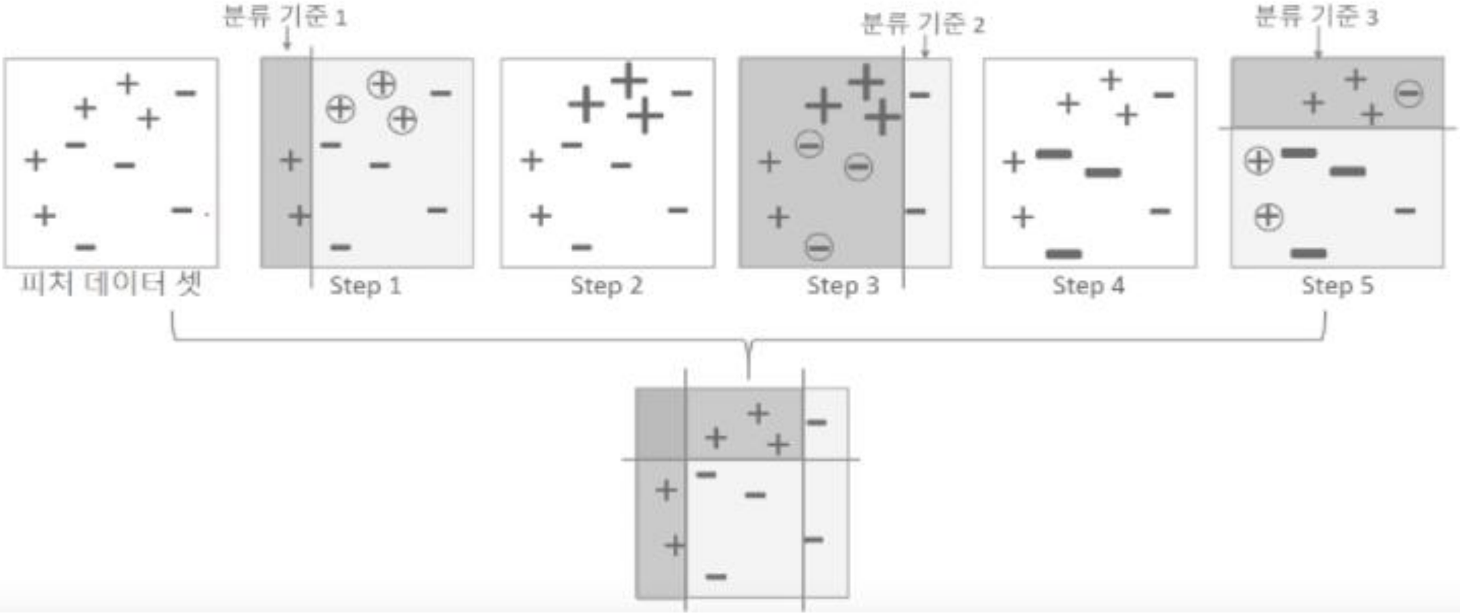
04. XG Boost VS Light GBM

01. 사전지식

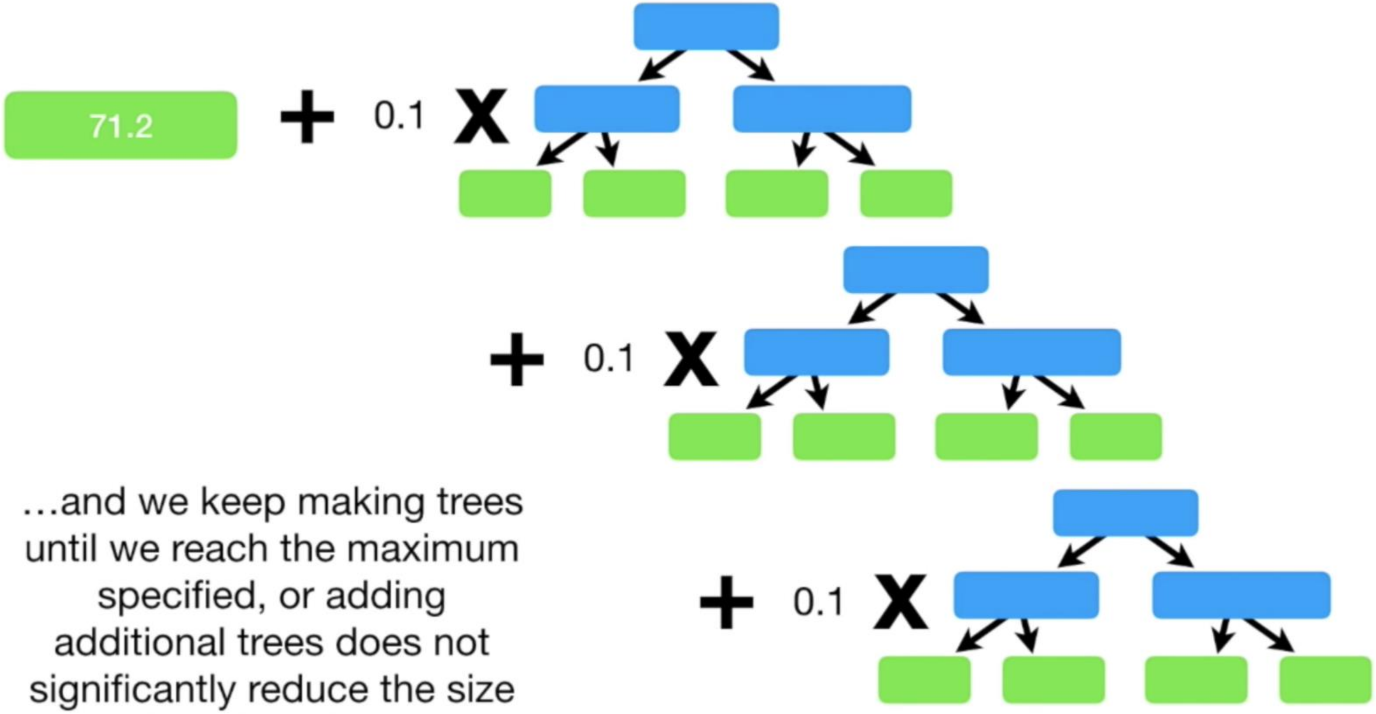
앙상블 학습



Boosting?



GBM



…and we keep making trees until we reach the maximum specified, or adding additional trees does not significantly reduce the size of the **Residuals**.

…etc…etc…etc…

02. XGBoost

XGBoost

- 과적합 규제
- CPU병렬처리 지원

- Y = $M(x) + error(1)$

XGBoost

- $\mathrm{Y} = M(x) + error(1)$
- $error(1) = G(x) + error(2)$

- $\mathrm{Y} = M(x) + error(1)$
- $error(1) = G(x) + error(2)$
- $error(2) = H(x) + error(3)$

- $\mathrm{Y} = M(x) + error(1)$
- $error(1) = G(x) + error(2)$
- $error(2) = H(x) + error(3)$

- $\mathrm{Y} = M(x) + G(x) + H(x) + error(4)$

- Y = $M(x)\ +\ error(1)$
- $error(1) = G(x)\ +\ error(2)$
- $error(2) = H(x) + error(3)$

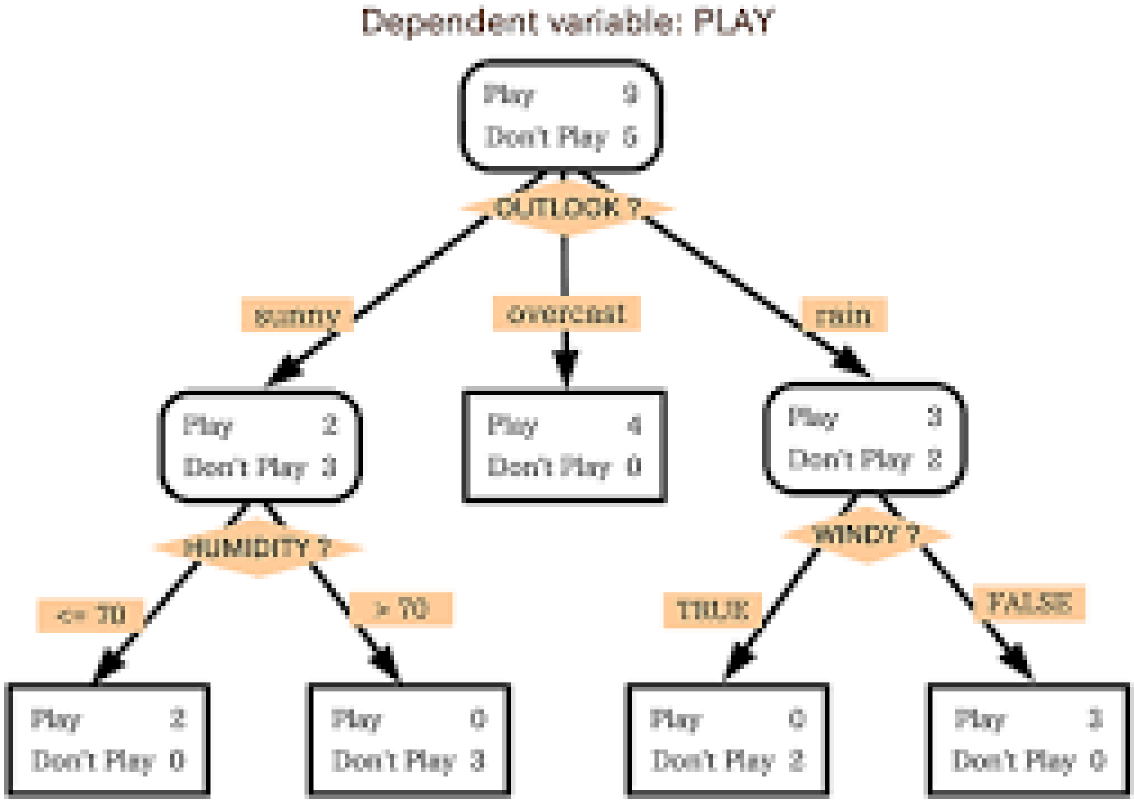
- Y = $M(x) + G(x) + H(x) + error(4)$

- Y = $\alpha M(x) + \beta G(x) + rH(x) + error(4)$

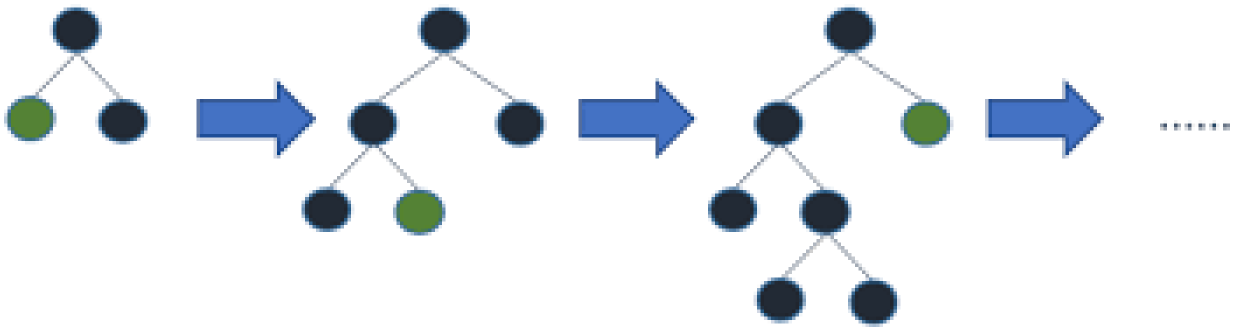
03. LightGBM

LightGBM

- 과적합에 민감합니다
- CPU병렬처리 지원 + GPU지원


Dependent variable: PLAY
Play 9
Don't Play 5
OUTLOOK ?
sunny
overcast
rain
Play 2
Don't Play 3
Play 4
Don't Play 0
Play 3
Don't Play 2
HUMIDITY ?
<= 70
> 70
WINDY ?
TRUE
FALSE
Play 2
Don't Play 0
Play 0
Don't Play 3
Play 0
Don't Play 2
Play 3
Don't Play 0

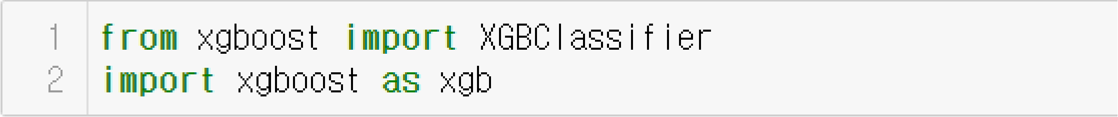
LightGBM



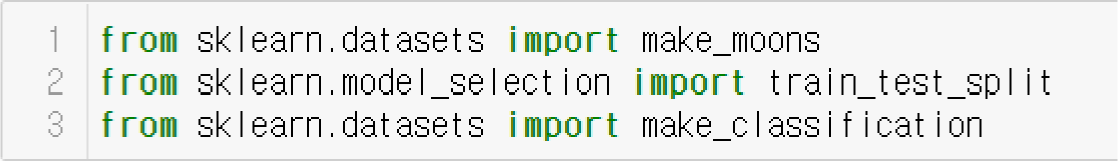
Leaf-wise tree growth

04. XGBoost VS LightGBM

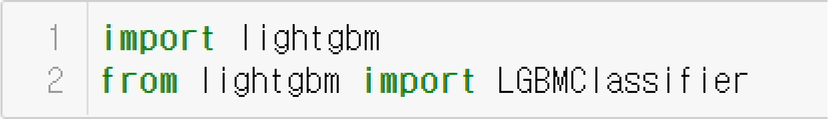
XGBoost VS LightGBM



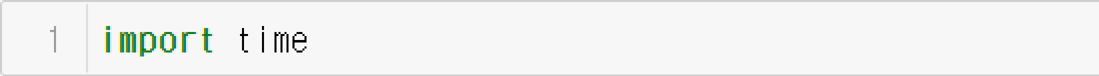
```python
from xgboost import XGBClassifier
import xgboost as xgb
```



```python
from sklearn.datasets import make_moons
from sklearn.model_selection import train_test_split
from sklearn.datasets import make_classification
```

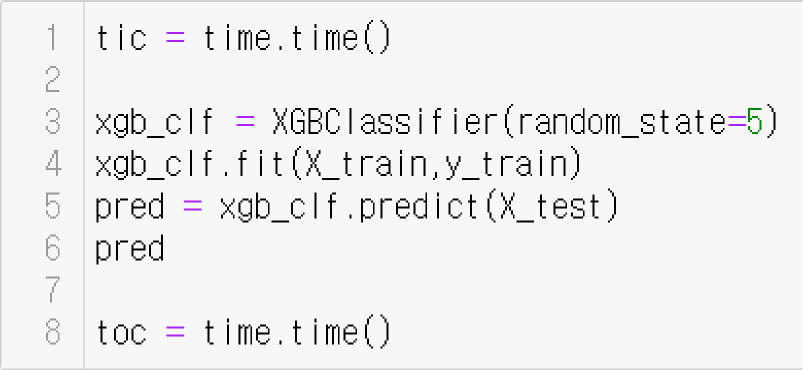


```python
import lightgbm
from lightgbm import LGBMClassifier
```

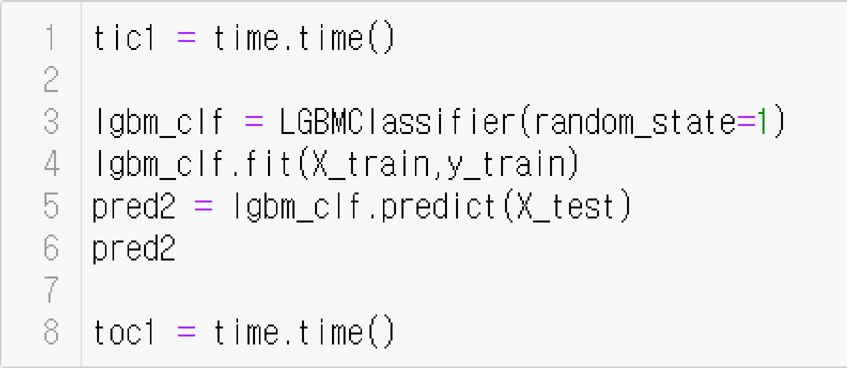


```python
import time
```

XGBoost VS LightGBM

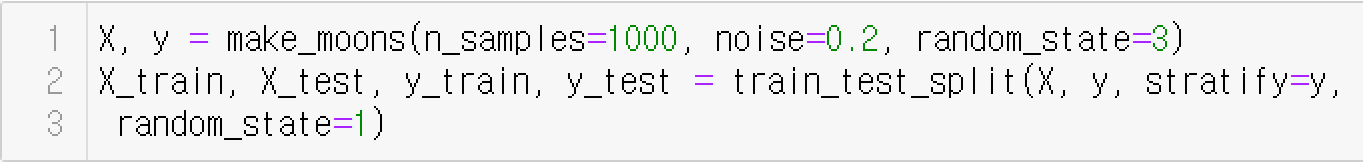


```
tic = time.time()

xgb_clf = XGBClassifier(random_state=5)
xgb_clf.fit(X_train,y_train)
pred = xgb_clf.predict(X_test)
pred

toc = time.time()
```

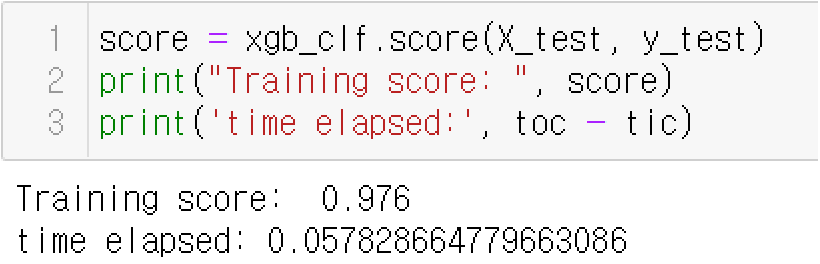


```
tic1 = time.time()

lgbm_clf = LGBMClassifier(random_state=1)
lgbm_clf.fit(X_train,y_train)
pred2 = lgbm_clf.predict(X_test)
pred2

toc1 = time.time()
```

XGBoost VS LightGBM

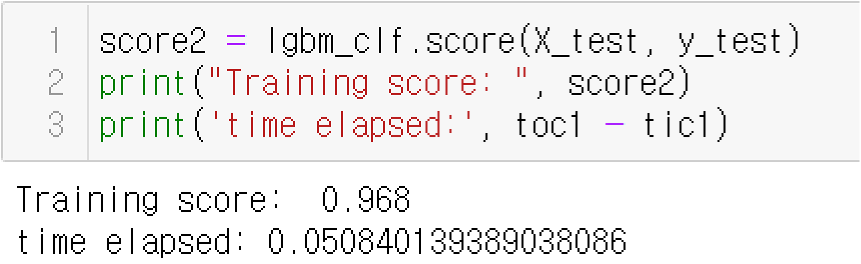


```
X, y = make_moons(n_samples=1000, noise=0.2, random_state=3)
X_train, X_test, y_train, y_test = train_test_split(X, y, stratify=y,
 random_state=1)
```



```
score = xgb_clf.score(X_test, y_test)
print("Training score: ", score)
print('time elapsed:', toc - tic)
```

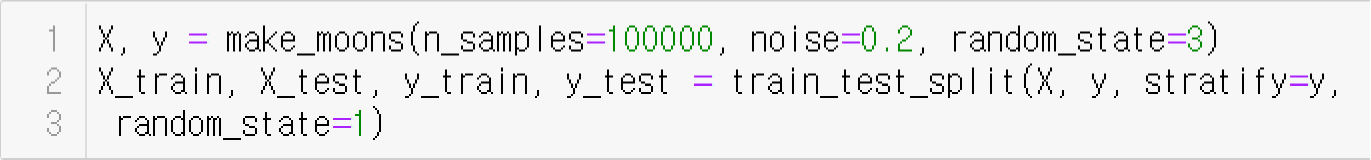
```
Training score:  0.976
time elapsed: 0.05782864779663086
```



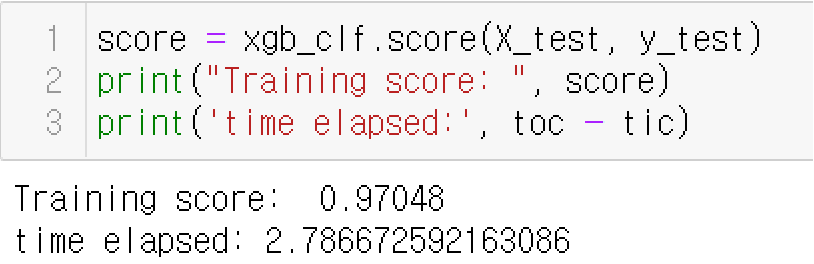
```
score2 = lgbm_clf.score(X_test, y_test)
print("Training score: ", score2)
print('time elapsed:', toc1 - tic1)
```

```
Training score:  0.968
time elapsed: 0.05084013938903808б
```

XGBoost VS LightGBM

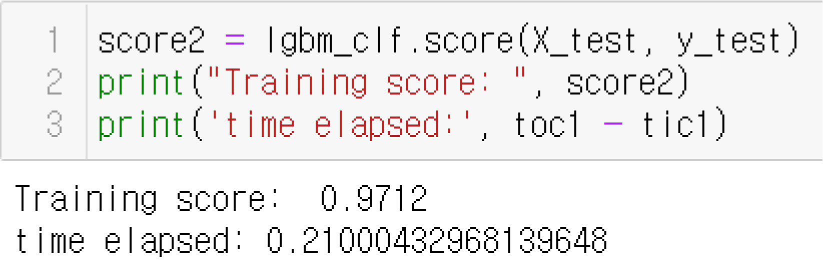


```
X, y = make_moons(n_samples=100000, noise=0.2, random_state=3)
X_train, X_test, y_train, y_test = train_test_split(X, y, stratify=y,
 random_state=1)
```



```
score = xgb_clf.score(X_test, y_test)
print("Training score: ", score)
print('time elapsed:', toc - tic)
```

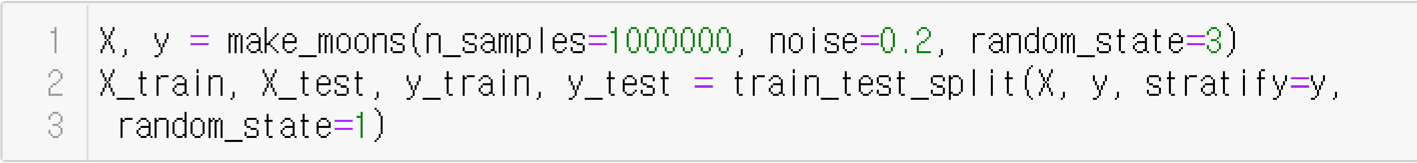
```
Training score:  0.97048
time elapsed: 2.786672592163086
```



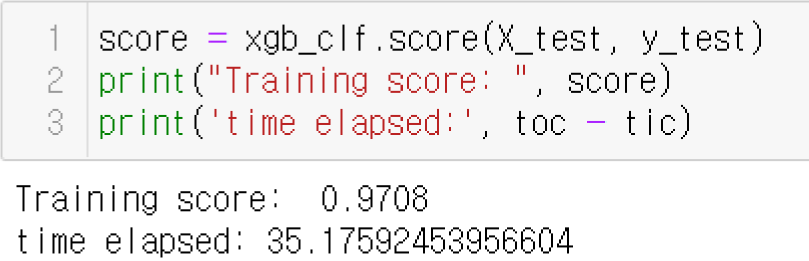
```
score2 = lgbm_clf.score(X_test, y_test)
print("Training score: ", score2)
print('time elapsed:', toc1 - tic1)
```

```
Training score:  0.9712
time elapsed: 0.21000432968139648
```

XGBoost VS LightGBM

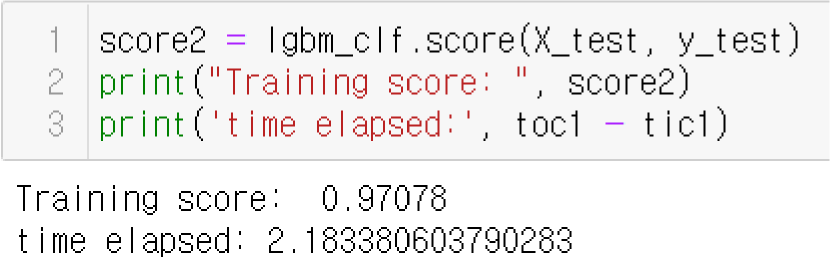


```python
X, y = make_moons(n_samples=1000000, noise=0.2, random_state=3)
X_train, X_test, y_train, y_test = train_test_split(X, y, stratify=y,
 random_state=1)
```



```python
score = xgb_clf.score(X_test, y_test)
print("Training score: ", score)
print('time elapsed:', toc - tic)
```

```
Training score:  0.9708
time elapsed: 35.17592453956604
```



```python
score2 = lgbm_clf.score(X_test, y_test)
print("Training score: ", score2)
print('time elapsed:', toc1 - tic1)
```

```
Training score:  0.97078
time elapsed: 2.183380603790283
```

Question

감사합니다☺