

앙상블 실습

GBM, LightGBM

GBM 실습 #1

- 데이터셋: 유방암 데이터
- 학습/시험 데이터: X, 학습/시험 데이터 라벨: Y

① 데이터 로드

```
1 from sklearn.datasets import load_breast_cancer
2 dataset = load_breast_cancer()
3 X = dataset.data
4 Y = dataset.target
```

② 데이터 분할 (학습/테스트)

```
1 from sklearn.model_selection import train_test_split
2 X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, stratify=Y, random_state=1 )
```

③ 모델 정의 및 학습

```
1 from sklearn.ensemble import GradientBoostingClassifier
2 clf = GradientBoostingClassifier(n_estimators=40)
3 clf.fit(X_train, Y_train)
```

GBM 실습 #1

- 데이터셋: 유방암 데이터
- 학습/시험 데이터: X, 학습/시험 데이터 라벨: Y

④ 학습 데이터 성능평가

```
1 from sklearn.metrics import accuracy_score
2 Y_train_pred = clf.predict(X_train)
3 accuracy_score(Y_train, Y_train_pred)
```

1.0

⑤ 테스트 데이터 성능평가

```
1 from sklearn.metrics import accuracy_score
2 Y_test_pred = clf.predict(X_test)
3 accuracy_score(Y_test, Y_test_pred)
```

0.9736842105263158

LightGBM 실습 #1

- 데이터셋: 유방암 데이터
- 학습/시험 데이터: X, 학습/시험 데이터 라벨: Y

① 데이터 로드

```
1 from sklearn.datasets import load_breast_cancer
2 dataset = load_breast_cancer()
3 X = dataset.data
4 Y = dataset.target
```

② 데이터 분할 (학습/테스트)

```
1 from sklearn.model_selection import train_test_split
2 X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2, stratify=Y, random_state=1 )
```

③ 모델 정의 및 학습

```
1 from lightgbm import LGBMClassifier
2 clf = LGBMClassifier(n_estimators=40)
3 clf.fit(X_train, Y_train)
```

LightGBM 실습 #1

- 데이터셋: 유방암 데이터
- 학습/시험 데이터: X, 학습/시험 데이터 라벨: Y

④ 학습 데이터 성능평가

```
1 from sklearn.metrics import accuracy_score
2 Y_train_pred = clf.predict(X_train)
3 accuracy_score(Y_train, Y_train_pred)
```

0.9978021978021978

⑤ 테스트 데이터 성능평가

```
1 from sklearn.metrics import accuracy_score
2 Y_test_pred = clf.predict(X_test)
3 accuracy_score(Y_test, Y_test_pred)
```

0.9736842105263158