

Bagging (4) - Wine 데이터셋 분류

이탈리아에서 생산되는 와인 3가지 종류의 성분 조사 데이터 셋

출처 : sklearn 예제 데이터셋

#01. 패키지 참조

진행되는 코드에 맞춰 필요한 패키지를 참조하세요.

#02. 데이터 가져오기

<https://data.hossam.kr/G02/wine.xlsx>

	alcohol	malic_acid	ash	alcalinity_of_ash	magnesium	total_phenols	flavanoids	nonflavonoids
0	14.23	1.71	2.43	15.6	127	2.80	3.06	0.28
1	13.20	1.78	2.14	11.2	100	2.65	2.76	0.26
2	13.16	2.36	2.67	18.6	101	2.80	3.24	0.30
3	14.37	1.95	2.50	16.8	113	3.85	3.49	0.24
4	13.24	2.59	2.87	21.0	118	2.80	2.69	0.39

#03. 데이터 전처리

1. 독립/종속 변수 분리

독립변수의 이름은 `class` 입니다.

```
((178, 13), (178,))
```

2. 독립변수에 대한 데이터 표준화

StandardScaler 사용

```
array([[ 1.51861254, -0.5622498,  0.23205254, -1.16959318,  1.91390522,
         0.80899739,  1.03481896, -0.65956311,  1.22488398,  0.25171685,
         0.36217728,  1.84791957,  1.01300893]])
```

3. 훈련 데이터와 검증 데이터 분리

```
((124, 13), (54, 13), (124,), (54,))
```

#04. 분류 모델중에서 가장 적합한 알고리즘 찾기

singleML 함수를 사용하여 가장 적합한 알고리즘을 선별합니다.

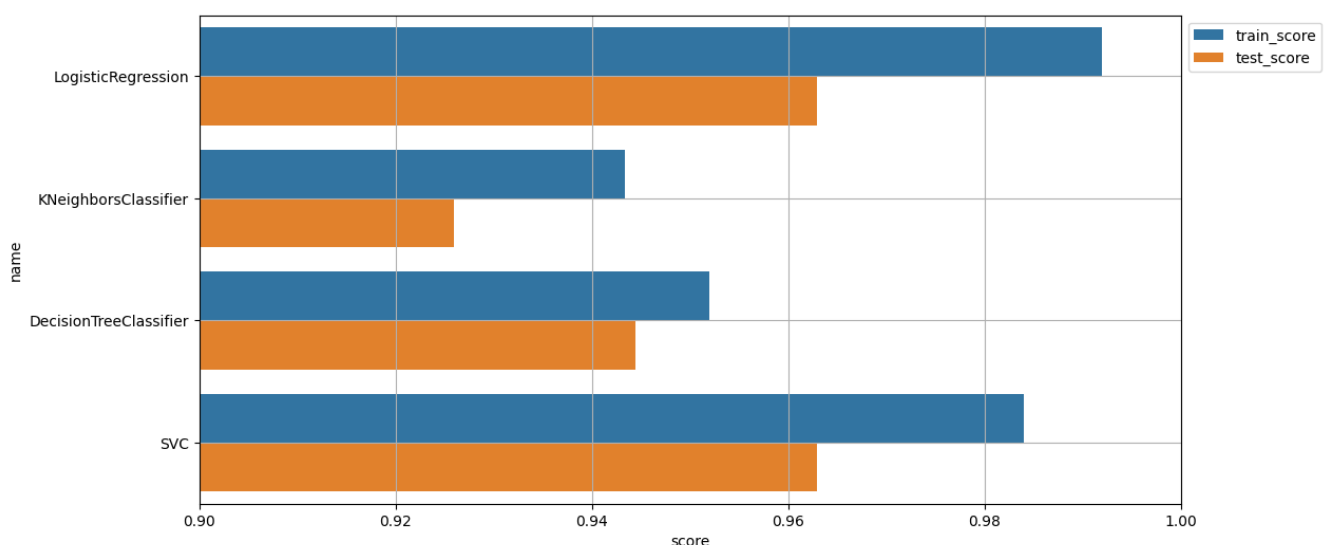
1. 알고리즘 수행 함수 정의 (singleML)

2. 적합한 알고리즘 찾기

테스트할 알고리즘 리스트를 정의한 후 각 알고리즘의 스코어를 데이터프레임으로 생성합니다.

	name	train_score	test_score
0	LogisticRegression	0.992000	0.962963
1	KNeighborsClassifier	0.943333	0.925926
2	DecisionTreeClassifier	0.952000	0.944444
3	SVC	0.984000	0.962963

3. 결과표의 재배치 및 알고리즘 별 성능 비교 시각화



#05. 선정된 알고리즘을 기반으로 Bagging에 대한 최적 파라미터 찾기

1. Bagging 모델 구현

```
{'bootstrap': False, 'bootstrap_features': True, 'n_estimators': 30}
```

	bootstrap	bootstrap_features	n_estimators	mean_test_score
4	False	True	30	1.000
5	False	True	50	1.000
0	True	True	30	0.992
1	True	True	50	0.992
6	False	False	30	0.992
7	False	False	50	0.992
2	True	False	30	0.984
3	True	False	50	0.984

2. 정확도 확인

1.0

3. 최적의 파라미터를 갖는 객체를 활용하여 예측결과 생성후 정확도 확인

GridSearchCV 분류기 정확도: 0.9815