

기계학습 코딩 시험

문제: 와인 샘플 178개의 화학성분을 나타내는 특성 13개로 구성된 wine.data를 이용하여, 와인 포도 품종 3개 중 어느 것인가를 구분하는 분류기를 구성한다.

	Class label	Alcohol	Malic acid	Ash	Alcalinity of ash	Magnesium	Total phenols	Flavanoids	Nonflavanoid phenols	Proanthocyanis	Color intensity	Hue	OD280/OD315 of diluted wines	Proline
0	1	14.23	1.71	2.43	15.6	127	2.80	3.06	0.28	2.29	5.64	1.04	3.92	1065
1	1	13.20	1.78	2.14	11.2	100	2.65	2.76	0.26	1.28	4.38	1.05	3.40	1050
2	1	13.16	2.36	2.67	18.6	101	2.80	3.24	0.30	2.81	5.68	1.03	3.17	1185
3	1	14.37	1.95	2.50	16.8	113	3.85	3.49	0.24	2.18	7.80	0.86	3.45	1480
4	1	13.24	2.59	2.87	21.0	118	2.80	2.69	0.39	1.82	4.32	1.04	2.93	735

1. 데이터 준비

- e강의동(<http://lms.sunmoon.ac.kr>)에서 wine.data, wine.names.txt를 다운로드 받는다.
(또는, <https://archive.ics.uci.edu/ml/datasets/wine> 에서 직접 다운로드한다.)
- wine.names.txt에는 데이터에 대한 설명이 들어 있다.

2. pandas를 이용하여 데이터를 읽어 들이고, class label과 특성 13개를 나눈다.

3. train_test_split을 이용하여, 학습 데이터와 테스트 데이터로 나눈다. 'stratify=y'로 하여 클래스 별 데이터 비율이 유지되도록 한다.

4. 적절한 분류기를 선정하여, 학습데이터로 학습시키고, 테스트 데이터에 대한 score를 출력한다. 분류기에 따라서, 데이터에 대한 스케일 변환을 실시하고, 학습시킨다.

5. 출력에 학과, 학번, 이름, 분류기명, 테스트 score가 나오도록 코딩한다. score는 저장된 값이거나, 직접 호출된 값이어야 한다. (즉, 숫자를 입력하면 안 됨)

6. 다음과 같은 하나의 샘플에 대하여 분류 결과가 출력되도록 하시오. 즉, predict()를 사용하고, 필요하다면 스케일링도 해야 한다.

```
X_new = np.array([[13.24, 2.59, 2.87, 21.0, 118, 2.80, 269, 0.39, 1.82, 4.32, 1.04, 2.93, 735]])
```

7. 소스코드와 결과 화면 캡처를 첨부하여 제출한다.