

2025-2 이미지딤러닝 Homework

- ▶ 제출기한 : 10월 24일 23:39 (지각시 시간당 -1점)
- ▶ 각 문제에 대한 답을 작성한 파이썬 파일 “HW1_학번.py”(1개)와 출력 캡처 파일(.png or .jpg) 1개를 LMS 과제함에 제출
- ▶ 주석으로 문제번호 작성
- ▶ 처음 부분에는 주석으로 학번/이름 작성
- ▶ for문, if문 등을 사용하는 것이 아닌 수업시간에 배운 함수 및 연산식을 사용하여 코드를 작성

보충코드

```
import pandas as pd

df = pd.DataFrame({'Weight':[45, 88, 56, 15, 71],
                  'Age':[14, 25, 55, 8, 21]})

print(df) #판다스 데이터 프레임
print(df.values) #넘파이 배열
print(df['Weight'].values) #넘파이 배열
```

- ▶ 강의 자료 및 보충 코드 외 명령어 사용시 횟수 당 -10점
- ▶ 상수형 숫자 사용시 감점(예. 평균 계산시 90/40)
- ▶ 데이터 파일의 경로는 “C:/cv_workspace”하위에 있어야함(경로 틀리면 감점)
예. #dir=“/root/~” #주석으로 경로 처리
dir=“C:/ml_workspace~”

* 의류 이미지 데이터 28,000개를 활용하여 이를 다양한 머신러닝 알고리즘에 학습시키고 의류 종류를 예측하며 성능을 비교하고자 한다.

* 업로드한 “data.csv” 파일을 활용하여 문제 해당하는 코드를 작성한다.

열 이름	열 설명	속성	행 번호
px0 ~ px783	각 이미지의 픽셀 값 (28×28 = 784개 픽셀)	정수형 0~255	
label	의류 종류를 나타내는 분류 라벨	범주형	0 ~ 6999
			7000 ~ 13999
			14000 ~ 20999
			21000 ~ 27999

(3점) 1. 파일을 읽고 파일의 행과 열의 개수를 출력하는 코드를 작성하시오.

실행결과

```
(28000, 785)
```

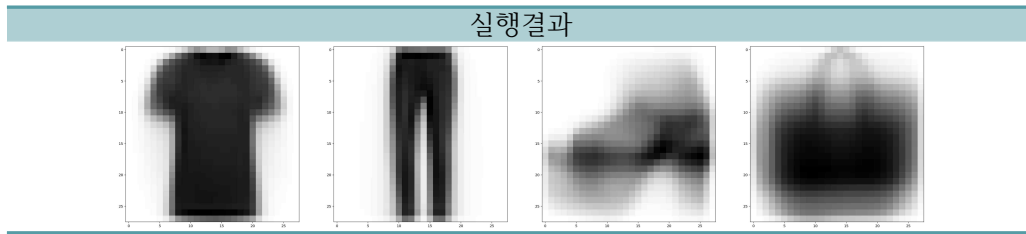
(7점) 2. 학습용 입력 데이터와 정답 데이터를 분리하고 각 데이터의 행과 열 개수를 출력하되 정답 데이터의 경우, 라벨 종류와 각 클래스별 데이터 개수를 함께 출력하는 코드를 작성하시오.

실행결과

```
X: (28000, 784)
y: (28000,)
(array([0, 1, 5, 8]), array([7000, 7000, 7000, 7000]))
```

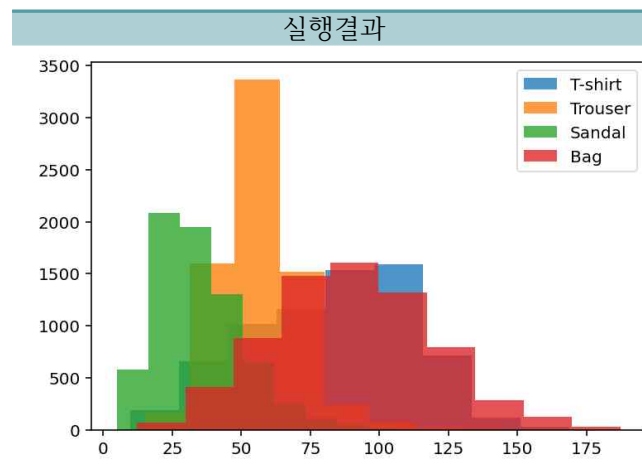
(13점) 3. 각 라벨별로 모든 이미지의 픽셀값 평균을 계산하고 이를 28×28 형태로 아래와 같이 출력하는 코드를 작성하시오.

- figure 사이즈: 가로 40, 세로 10
- for문 사용 가능



(12점) 4. 각 라벨별로 픽셀 평균값 분포 히스토그램을 아래와 같이 출력하는 코드를 작성하시오.

- for문 사용 가능



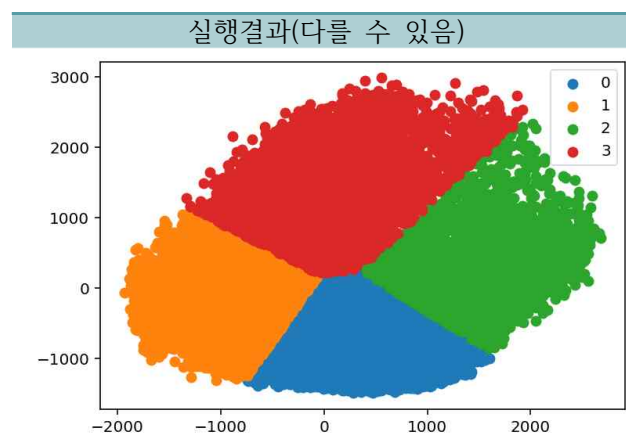
(10점) 5-1. PCA 주성분 분석으로 학습 데이터를 2개의 열로 축소한 후, 4개의 군집으로 분류하는 k-평균 알고리즘으로 학습시키고 예측된 라벨별 데이터 개수를 출력하는 코드를 작성하시오.

실행결과(다를 수 있음)

```
(array([0, 1, 2, 3], dtype=int32), array([8807, 8700, 4613, 5880]))
```

(7점) 5-2. 위 코드에서 예측한 라벨에 해당하는 데이터를 산점도로 시각화 하는 코드를 작성하시오.

- for문 사용 가능



(10점) 6. PCA 주성분 분석으로 30개의 열로 축소한 후, 테스트 데이터는 전체 데이터의 22%로 사용하여 랜덤하게 훈련 데이터와 테스트 데이터를 분리하고 각 데이터의 행과 열 개수를 출력하는 코드를 작성하시오.

실행결과
훈련 데이터 X: (21840, 30)
훈련 데이터 y: (21840,)
테스트 데이터 X: (6160, 30)
테스트 데이터 y: (6160,)

(10점) 7. 학습 시킬 데이터를 모두 StandardScaler로 정규화하고 일부를 출력하는 코드를 작성하시오.

실행결과
훈련 데이터 X 정규화 [[0.39844136 -1.19690862 -0.76823836 -0.5055324 0.23854425 0.26314312생략... 0.42827638 -1.75337014 0.23280841 -1.40834924]]
테스트 데이터 X 정규화 [[-1.56465423e+00 -2.01430248e-01 3.15075158e-01 4.90651106e-01생략... 2.80819754e-01 7.01165650e-01 5.97728327e-01 -7.50838473e-01]]

(10점) 8. 로지스틱 손실함수를 사용한 경사 하강법 모델을 정의하고 정규화된 학습 데이터를 이용하여 모델을 학습시킨 뒤 훈련 데이터와 테스트 데이터의 점수를 각각 출력하는 코드를 작성하시오. 또한 7개의 폴드 교차 검증을 수행하고 테스트 데이터 평균 점수를 아래와 같이 출력하는 코드를 작성하시오.

실행결과(다를 수 있음)
[로지스틱]훈련 데이터 점수: 0.9736263736263736
[로지스틱]테스트 데이터 점수: 0.9696428571428571
[교차검증]테스트 데이터 점수: 0.9729395604395605

(10점) 9. 결정트리 모델을 정의하고 정규화된 학습 데이터를 이용하여 모델을 학습시킨 뒤 훈련 데이터와 테스트 데이터의 점수를 각각 출력하는 코드를 작성하시오. 또한 7개의 폴드 교차 검증을 수행하고 테스트 데이터 평균 점수를 아래와 같이 출력하는 코드를 작성하시오.

실행결과(다를 수 있음)
[결정트리]훈련 데이터 점수: 1.0
[결정트리]테스트 데이터 점수: 0.9599025974025974
[교차검증]테스트 데이터 점수: 0.9562728937728939

(8점) 10. 테스트 데이터 중 10개 데이터를 선택하여 위에서 학습된 모델로 예측한 값과 실제 값을 함께 출력하는 코드를 작성하시오.

실행결과(다를 수 있음)
[로지스틱]예측한 값: [5 0 5 8 8 0 5 0 8 0]
[결정트리]예측한 값: [5 0 5 8 8 0 5 0 8 0]
실제 정답: [5 0 5 8 8 0 5 0 8 0]