

BOAZ Spring Study : Weekly Assignment  
(Due date : 2023.07.24)

1. 다음은 어떤 데이터의 3가지 컬럼과 y 값이 담겨 있다. 3가지 컬럼에 적절한 가중치를 곱하여 y를 만들어야 할 때, 아래 물음에 답하세요. (30 points)

Index	Column 1	Column 2	Column 3	y
1	1	3	5	2
2	1	1	0	3
3	1	1	2	-1
4	1	3	3	2

(1) 정규방정식을 이용하여 최소제곱해를 구하세요.

(2) (2)에서 구한 최소제곱해를 이용하여  $\|b - Ax\|$ 를 구하고, 이것이 의미하는 바가 무엇인 지 간단하게 설명하세요.

< Python Coding Assignment - 모든 문제는 Numpy를 이용하세요 >

2. 아래 순서를 따라 코드를 완성해 보세요. (30 points)

- 모든 값이 0인 100 x 2 행렬 X를 만들고, 첫 번째 열에는 10부터 100까지의 자연수 중에서 랜덤으로 100개를, 두 번째 열에는 100부터 2000까지의 자연수 중 랜덤으로 100개를 저장한다.
- 행렬 X와 동일한 크기(100 x 2)를 갖는 행렬을 standard\_x라는 변수로 만든다.
- 처음에 만들었던 행렬 X에서 열 단위로 평균이 0이고 표준편차가 1인 표준정규분포로 표준화하고, 표준화된 결과를 standard\_x 변수에 각각 저장한다.
- `print(standard_x[:, 0].mean(), standard_x[:, 1].mean(), standard_x[:, 0].std(), standard_x[:, 1].std())`를 출력하여 각각 출력이 0, 0, 1, 1에 가까운 지 확인한다.

< -- 노선에 첨부된 population.csv를 불러온 상태에서 시작하십시오 -- >

3. 해당 데이터 파일인 population.csv를 불러오고, 전체 데이터에 대해서 아래와 같은 분석을 수행하는 프로그램을 Numpy ndarray의 indexing and slicing 기법을 이용하여 작성하세요. 데이터 설명은 아래와 같습니다. (40 points)

< 데이터 설명 >

데이터는 총 3개의 열로 구성되어 있으며, 순서대로 성별, 나이, 교육정도코드입니다.  
각 코드별 내용은 아래와 같습니다.

성별 : 1(남자), 2(여자)

나이 : 0~85 사이의 숫자

교육정도(학력) : 1(미취학), 2(초등학생), 3(중학생), 4(고등학생), 5(4년제 미만 대학),  
6(4년제 이상 대학), 7(대학원 석사 과정), 8(대학원 박사 과정)

시작 전, 아래와 같은 코드를 추가하여 popluation.csv를 numpy의 ndarray  
형태로 읽어올 수 있게 하세요.

```
import numpy as np
data_arr = np.loadtxt("population.csv", delimiter=",", dtype=np.int64)
print("[Data Loaded]:", data_arr.shape, '\n')
## 출력 결과 : (463447, 3)
```

(1) 연령대별 남녀 성비를 출력하세요. 연령대는 10대 미만(0~9), 10대(10~19), 20대(20~29)와  
같이 구분.) 출력 예시는 아래와 같습니다.

< 출력 결과 예시 >

	남성	여성
20대미만	0.35	0.35
20대	0.25	0.25
30대	0.20	0.20
40대	0.15	0.15
50대이상	0.35	0.35

(2) 연령대별/성별( 및 전체) 대학 진학률을 출력하세요. 연령대는 20대, 30대, 40대, 50대  
이상으로만 분류하고, 대학 진학 여부는 “대학(4년제 미만)” 코드인 5 이상이면 대학 진학  
으로 판단하세요.

< 출력 결과 예시 >

	남성	여성	전체
20대	0.55	0.57	0.55
30대	0.34	0.56	0.44
40대	0.58	0.53	0.55
50대 이상	0.33	0.44	0.55

[제출 방법]

1. 1번 문항 관련 풀이 (선택사항)
2. 2, 3번 문항 관련된 ipynb 파일 (하나의 ipynb 파일 안에 두 문제를 풀어주세요.)