

**BOAZ Spring Study : Weekly Assignment**  
(Due date : 2023.03.26)

[필기]

1. 아래 함수의 도함수를 구하세요. (10 points)

(1)  $y = e^x \sin 2x$

(2)  $y = \frac{x^2}{x+1}$

2.  $f(x, y, z) = x^2 + 2xy + yz^3$ 에 대해  $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}$ 를 구하고, (1, 1, 2)에서의 편미분계수를 구하시오. (5 points)

3. 아래 물음에 답하시오. (10 points)

(1)  $z = x^2 + y \cos(x+y)$ 이고,  $x = t^2 - t$ ,  $y = \frac{\pi}{2}t^3$ 일 때,  $\frac{dz}{dt}$ 를 구하시오.

(2)  $z(x, y) = x^2y^2$ ,  $x = u^2 + v^2$ ,  $y = 2uv$ 일 때,  $\frac{\partial z}{\partial u}, \frac{\partial z}{\partial v}$ 를 모두 구하시오.

4. 함수  $f(x) = x^3 - 3x^2 - 9x + 1$ 에서 local minimum과 local maximum을 찾으시오. (10 points)

5. 아래 데이터는 BOAZ 20기 사람들의 통학 거리에 따른 소요 시간과 관련된 데이터이다. 선형 회귀를 이용하여 데이터를 설명하려고 할 때, Gradient Descent Algorithm을 이용하여 최적의 파라미터를 찾으려고 한다. 두 번째 예폭을 지난 후의 선형 회귀 방정식을 쓰시오. (단, learning rate = 0.1, 파라미터의 초깃값은  $a = 10$ ,  $b = 1$ , Loss function은 MSE로 가정.) (15 points)

Index	거리(km)	소요 시간(min)
1	1.5	15
2	4	37
3	3	32
4	5	48

[Python Coding - 실습 파일 참고]

< -- 실습 과제 시작 전, Notion의 train.csv를 다운 받아 적절한 경로에 저장하세요 -- >

6. 다음은 Python을 이용하여 Gradient Descent를 구현해보려고 한다. (25 points)

(1) 아래의 순서대로 Gradient Descent를 구현해보세요. (12 points)

- a. 평균이 0이고, 표준편차가 1인 정규분포를 따르는 집단에서 2개를 추출하여 array를 생성하고, array의 첫 번째 요소를  $y = ax + b$ 에서의 a로, 두 번째 요소를 b로 저장한다. (np.random.randn이나 np.random.normal을 적절히 사용하면 됩니다.)
- b. Gradient Descent를 실행하기 전, loss function을 계산해보세요. (error에다가 채우시면 됩니다.)
- c. Gradient Descent를 실행하기 위해 learning rate를 0.05, epoch를 100으로 설정하세요.
- d. Loss Function이 MSE라는 것을 이용해서 반복문 안에 A의 gradient, B의 gradient를 계산하고, A와 B를 update하는 코드를 작성하세요.

(2) (1) 실행 후에 A, B는 얼마이고, 그 때의 loss 값은 얼마인가? (4 points)

(3) (1)의 epoch이 50, 100, 500, 1000인 경우의 loss 값을 비교하세요. (4 points)

(4) (3)에서 500인 경우와 1000의 경우가 거의 비슷하게 나오는데, 이유를 설명해 보세요. (5 points)

7. 아래는 Pandas를 이용해서 데이터 전처리를 하는 것과 관련한 문항입니다. 물음에 답하세요. (25 points)

(1) df 데이터에서 ProdTaken 컬럼이 어떤 값으로 이루어져 있는 지 코드를 쓰고, 각각 몇 개씩 있는 지 코드를 쓰세요. (3 points)

(2) df 데이터에서 어떤 컬럼이 결측값이 존재하는 지 쓰고, 각 컬럼이 몇 개씩 결측값을 가지는 지 코드를 작성하세요. (3 points)

(3) df 데이터에서 ProdTaken 별로 NumberOfTrips의 평균, 표준편차, 최대, 최솟값을 출력하는 코드를 쓰세요. (3 points)

(4) df 데이터에서 Age 열과 TypeOfContact에 존재하는 결측값을 채우려고 합니다.. 이때, Age는 중앙값(median)으로, TypeofContact는 최빈값으로 대체하는 코드를 작성하세요. (5 points)

(5) 아래와 같은 규칙에 맞게 파생변수를 생성하려고 한다. 아래 규칙에 맞게 세 개의 파생변수를 생성하세요. (6 points)

- DurationOfPitch, NumberOfFollowups, PitchSatisfactionScore를 모두 더한 값을 Service라는 새로운 파생변수로 생성
- ProductPitched가 Basic이면 1, Standard면 2, Deluxe면 3, Super Deluxe면 4, King이면 5로 저장하는 convert\_present 파생변수 생성
- Age를 5개의 구간으로 나누어서 Age\_convert라는 새로운 파생변수 생성(cut 함수 이용)

(6) convert\_present가 3보다 크고, Age가 40보다 작은 사람들만 새롭게 추출해서 filtered\_data라는 변수에 저장하고, index를 초기화하세요. (5 points)