

# 기계학습시스템설계 (ELEC 0504)

## 과제 01

2장. 머신러닝을 위한 기초지식

3장. Numpy, Pandas

# 과제 평가 방식

- 매 주 과제당 총점 100점 기준
  - 10점: 코드의 맨 처음 줄은 본인의 학번 및 이름을 주석으로 작성  
주석은 # 을 줄 맨 앞에 삽입 ex) # 20221234 홍길동
  - 30점: 코드 가독성, 코드가 이해가능 하도록 작성되었는가?
  - 30점 : 작성한 코드의 실행 가능 여부
  - 30점 : 작성한 코드 실행 후 출력값 일치 여부

## **\*\* 주의 사항 \*\***

- 제출기한: 매주 화요일 자정(밤12시)까지 LMS에 업로드
- 지각 제출시 1시간당 -10점

# 과제 제출 방법

- 평가 환경
  - Python 3.9.18
  - Spyder 5.4.3
- 파일명: 각 문제에 명시
- 제출방법
  - 코드파일(.py)들과 스크린샷파일(.png)들을 압축하여 1개의 zip파일로 제출

**압축파일명: hw01.zip**

# 과제 01-1

소스코드 파일명: 01-1.py

스크린샷 파일명: 01-1.png

- numpy 라이브러리를 이용하여 1에서 100사이의 중복되지 않은 임의의 정수 원소 N개를 포함하는 행벡터(row vector)를 생성하고, 이 행벡터의 L1 Norm 과 L2 Norm을 계산하는 함수를 구현하라. 또한 함수의 결과값을 numpy 라이브러리에서 제공하는 L1 Norm과 L2 Norm 를 계산하는 함수의 결과값과 비교하라.

$$\|\mathbf{x}\|_p = \left( \sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}$$

```
N을 입력하세요 >> 10
arr : [69 59 77 55 32 82 99 49 21 11]

내가 만든 L1-Norm : 554.0
numpy의 L1-Norm : 554.0

내가 만든 L2-Norm : 194.18547834480313
numpy의 L2-Norm : 194.18547834480313
```

# 과제 01-2

소스코드 파일명: 01-2.py

스크린샷 파일명: 01-2.png

- 1) numpy 라이브러리를 이용하여 행렬  $A \in \mathbb{Z}^{k \times k}$ 와  $B \in \mathbb{Z}^{k \times k}$ 의 hadamard product 를 계산하는 코드를 구현하고 이를 numpy의 hadamard product 함수의 결과값과 비교하라 (k를 입력받고 원소의 범위는 1부터 10사이).
- 2) 또한,  $A \in \mathbb{Z}^{n \times m}$ 와  $B \in \mathbb{Z}^{m \times n}$ 의 dot product를 계산하는 코드를 구현하고 이를 numpy의 dot product 함수의 결과값과 비교하라 (n과 m은 입력받고 원소의 범위는 1부터 10사이).

```
행렬의 k를 입력하세요 >> 3
행렬 A
[[3 3 9]
 [7 5 2]
 [5 7 2]]
행렬 B
[[5 2 5]
 [7 8 5]
 [8 3 9]]
내가 만든 hadamard product
[[15.  6. 45.]
 [49. 40. 10.]
 [40. 21. 18.]]
numpy의 hadamard product
[[15  6 45]
 [49 40 10]
 [40 21 18]]
```

```
행렬의 n, m을 입력하세요 >> 3 4
행렬 A
[[9 3 8 9]
 [3 8 8 2]
 [9 2 1 2]]
행렬 B
[[3 9 6]
 [1 7 4]
 [4 4 5]
 [4 9 9]]
내가 만든 dot product
[[ 98. 215. 187.]
 [ 57. 133. 108.]
 [ 41. 117.  85.]]
numpy의 dot product
[[ 98 215 187]
 [ 57 133 108]
 [ 41 117  85]]
```

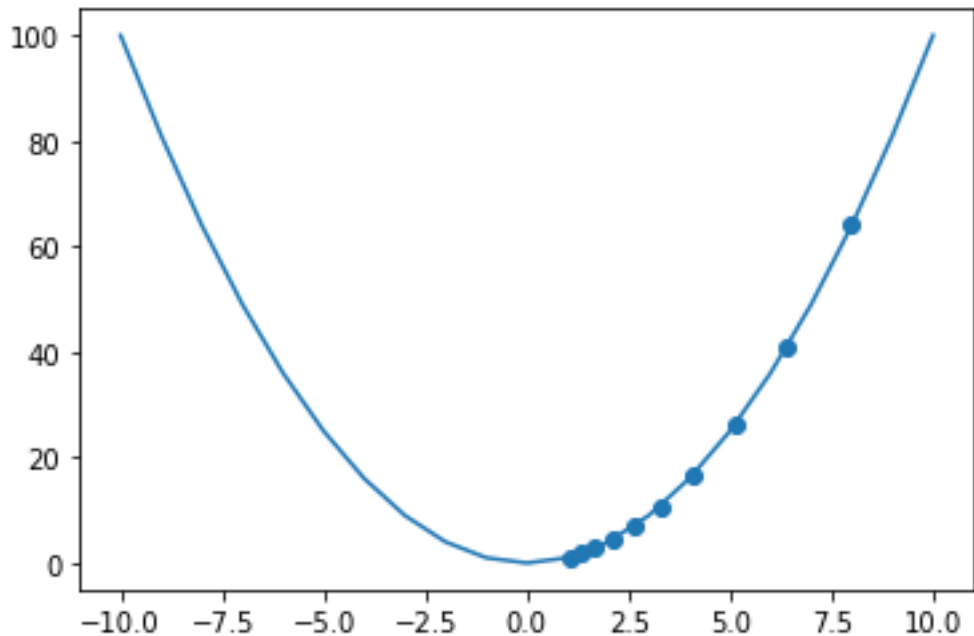
# 과제 01-3

소스코드 파일명: 01-3.py

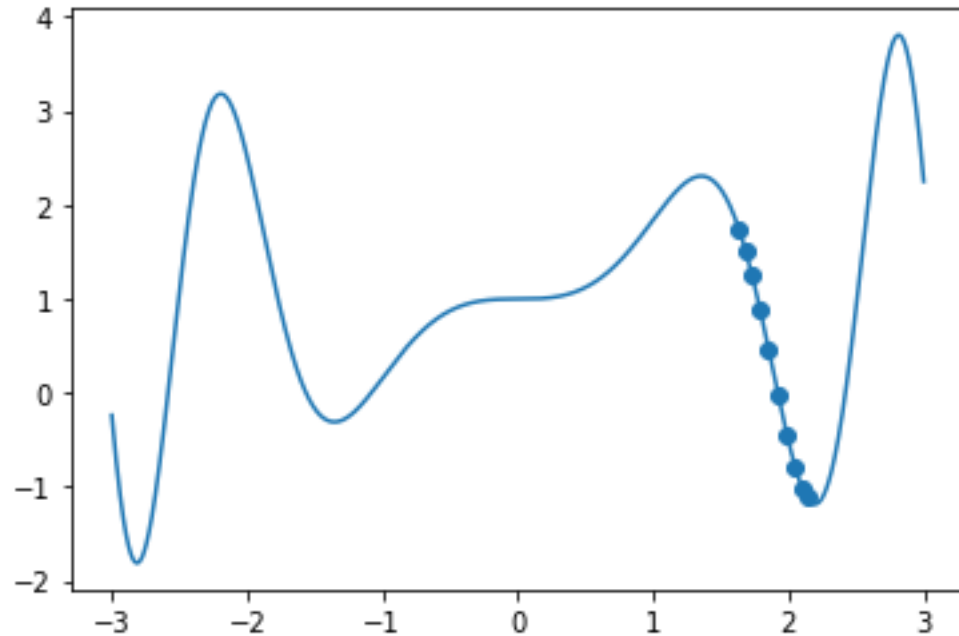
스크린샷 파일명: 01-3.png

- numpy와 matplotlib 라이브러리를 이용하여 아래 2가지 함수에 대해 경사 하강법을 이용하여 최소값을 찾는 코드를 구현하고 그래프를 그려라.

$f(x) = x^2, \tilde{x} = x - \eta \nabla f(x), x_0 = 10, \eta = 0.1$   
 $x = [-10, 10]$ , 간격 1 이고 반복은 10번



$f(x) = x \sin(x^2), \tilde{x} = x - \eta \nabla f(x), x_0 = 1.6, \eta = 0.01$   
 $x = [-3, 3]$ , 간격 0.01 이고 반복은 10번



# 과제 01-4

소스코드 파일명: 01-4.py

스크린샷 파일명: 01-4.png

```
data = {  
    'year':[2016, 2017, 2018],  
    'car': ['그랜저', '그랜저', '소나타'],  
    'name': ['홍길동', '고길동', '김둘리' ],  
    'number' : ['123하4567', '123허4567', '123호4567']  
}
```

```
In [231]: DF.head()  
Out[231]:  
   year  car name  number  
0  2016   그랜저  홍길동  123하4567  
1  2017   그랜저  고길동  123허4567  
2  2018   소나타  김둘리  123호4567
```

라는 dictionary를 만들고 이를 pandas의 dataframe으로 변환

## 과제 01-4

(1) DF를 보면 index 0 1 2 에 data가 있습니다. index 3에 "2017, 일론 , 테슬라 , 987하6543" 라는 새로운 정보를 추가하는 코드를 작성하시오.

	year	car	name	number
0	2016.0	그랜저	홍길동	123하4567
1	2017.0	그랜저	고길동	123허4567
2	2018.0	소나타	김둘리	123호4567
3	2017.0	테슬라	일론	987하6543

(2) 전체 데이터 중에서 year, car, number만 출력하시오.

	year	car	number
0	2016.0	그랜저	123하4567
1	2017.0	그랜저	123허4567
2	2018.0	소나타	123호4567
3	2017.0	테슬라	987하6543

(3) 2018년 이전의 차량과 관련된 정보만 출력하시오.

	year	car	name	number
0	2016.0	그랜저	홍길동	123하4567
1	2017.0	그랜저	고길동	123허4567
3	2017.0	테슬라	일론	987하6543



# 과제 01-5

소스코드 파일명: 01-5.py

스크린샷 파일명: 01-5.png

첨부된 csv데이터(2014년, 2015년, 2016년의 졸음운전 교통사고 데이터)를 이용하여 아래 문제에 대한 결과를 출력하는 코드를 작성하시오.

(1) 불러온 3개의 데이터를 병합하고 "구분"을 index로 설정하시오.

	사고(건)	사망(명)	부상(명)
구분			
2014년1월	169	7	347
2014년2월	145	3	268
2014년3월	206	12	358
2014년4월	182	4	360
2014년5월	218	17	474
2014년6월	233	13	457
2014년7월	229	14	431
2014년8월	222	9	447
2014년9월	225	15	430
2014년10월	210	14	392
2014년11월	209	10	386
2014년12월	178	12	329
2015년1월	205	8	375
2015년2월	175	8	375
2015년3월	237	4	492
2015년4월	227	10	486
2015년5월	231	9	504
2015년6월	207	8	401
2015년7월	273	7	556
2015년8월	268	19	534
2015년9월	252	10	501
2015년10월	214	14	475
2015년11월	199	3	398
2015년12월	213	8	428
2016년1월	192	5	387
2016년2월	174	6	328
2016년3월	217	7	435
2016년4월	216	7	419
2016년5월	239	13	522
2016년6월	200	12	362
2016년7월	227	9	460
2016년8월	230	7	490
2016년9월	187	13	347
2016년10월	183	10	367
2016년11월	200	5	418
2016년12월	168	4	364

# 과제 01-5

(2) 년도, 월별로 index를 설정(멀티인덱스) 하고 데이터를 보여 주시오.

년도	월	사고(건)	사망(명)	부상(명)
2014	1	169	7	347
	2	145	3	268
	3	206	12	358
	4	182	4	360
	5	218	17	474
	6	233	13	457
	7	229	14	431
	8	222	9	447
	9	225	15	430
	10	210	14	392
	11	209	10	386
	12	178	12	329
2015	1	205	8	375
	2	175	8	375
	3	237	4	492
	4	227	10	486
	5	231	9	504
	6	207	8	401
	7	273	7	556
	8	268	19	534
	9	252	10	501
	10	214	14	475
	11	199	3	398
	12	213	8	428
2016	1	192	5	387
	2	174	6	328
	3	217	7	435
	4	216	7	419
	5	239	13	522
	6	200	12	362
	7	227	9	460
	8	230	7	490
	9	187	13	347
	10	183	10	367
	11	200	5	418
	12	168	4	364

## 과제 01-5

(3) 년도 및 월별로 사망자를 출력하시오.

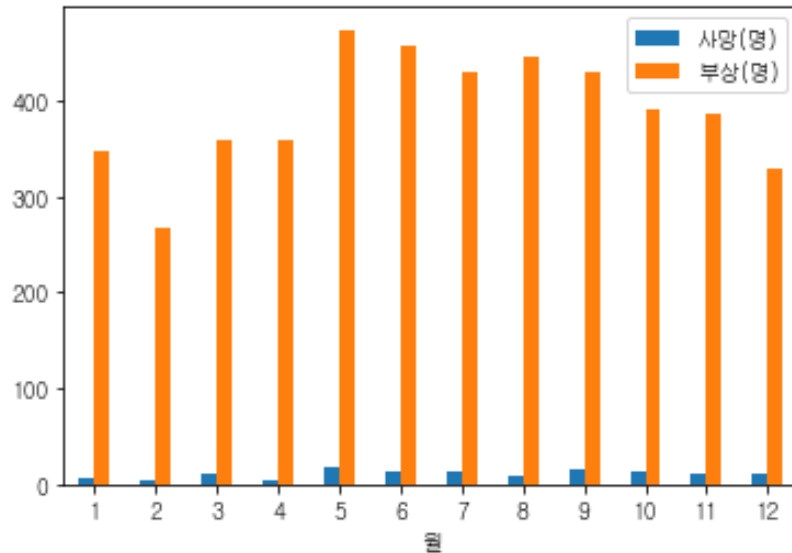
(4) 2016년 전체교통사고대비 사망률을 출력하시오.

사망(명)	
년도	
2014	10.833333
2015	9.000000
2016	8.166667
사망(명)	
월	
1	6.666667
2	5.666667
3	7.666667
4	7.000000
5	13.000000
6	11.000000
7	10.000000
8	11.666667
9	12.666667
10	12.666667
11	6.000000
12	8.000000

2016년 통계  
전체사고(건): 2433, 사망자(명):98, 사고대비사망율:4.03%

## 과제 01-5

(5) 2014년도 월별 사망, 부상 데이터를 bar차트로 보여주세요.



\* plot 출력시 한글 폰트 깨짐 문제 해결

```
from matplotlib import font_manager, rc
font_path = "C:/Windows/Fonts/NGULIM.TTF"
font = font_manager.FontProperties(fname=font_path).get_name()
rc('font', family=font)
```

(6) 2015년 대비 사망이 가장 많이 증가한 2016년도의 2개의 월을 구하세요.