

## MARG Python Study Assignment 3

### Question 1. 이미지 데이터 전처리

주어진 파일 “MARG Python Assignment 3 (Data).csv”는 오른쪽에 보이는 28X28 크기 흑백 구두 사진의 greyscale 수치 정보만을 담은 파일입니다.



일반적으로 이미지 파일은 각 픽셀의 color value 를 테이블 형태로 저장합니다. 이 파일의 경우 가로 28 픽셀, 세로 28 픽셀로 총 784 픽셀이 있고, 각 픽셀에 해당되는 색 정보를 0(흰색)~255(검은색) 사이의 정수 값으로 저장하고 있습니다. 정수 값은 CSV 형태로 저장되어 있습니다.

CSV 는 Comma Separated Values 의 약자로, 데이터를 저장하는 가장 간단한 형태의 파일 형식 중 하나입니다. 테이블의 형태를 띄고 있으며, 각 row 는 new line 으로, 각 column 은 쉼표로 구분합니다. 다음은 주어진 파일의 일부 내용입니다:

```
0,0,0,0,0,0,0,0,0,0,0,0,3,0,36,136,127,62,54,0,0,0,1,3,4,0,0,3
0,0,0,0,0,0,0,0,0,0,0,0,6,0,102,204,176,134,144,123,23,0,0,0,0,12,10,0
0,0,0,0,0,0,0,0,0,0,0,0,0,155,236,207,178,107,156,161,109,64,23,77,130,72,15
0,0,0,0,0,0,0,0,0,0,0,0,1,0,69,207,223,218,216,216,163,127,121,122,146,141,88,172,66
0,0,0,0,0,0,0,0,0,1,1,1,0,200,232,232,233,229,223,223,215,213,164,127,123,196,229,0
0,0,0,0,0,0,0,0,0,0,0,0,183,225,216,223,228,235,227,224,222,224,221,223,245,173,0
0,0,0,0,0,0,0,0,0,0,0,0,193,228,218,213,198,180,212,210,211,213,223,220,243,202,0
0,0,0,0,0,0,0,0,0,1,3,0,12,219,220,212,218,192,169,227,208,218,224,212,226,197,209,52
0,0,0,0,0,0,0,0,0,0,6,0,99,244,222,220,218,203,198,221,215,213,222,220,245,119,167,56
0,0,0,0,0,0,0,0,0,4,0,0,55,236,228,230,228,240,232,213,218,223,234,217,217,209,92,0
0,0,1,4,6,7,2,0,0,0,0,0,237,226,217,223,222,219,222,221,216,223,229,215,218,255,77,0
0,3,0,0,0,0,0,0,0,62,145,204,228,207,213,221,218,208,211,218,224,223,219,215,224,244,159,0
0,0,0,0,18,44,82,107,189,228,220,222,217,226,200,205,211,230,224,234,176,188,250,248,233,238,215,0
0,57,187,208,224,221,224,208,204,214,208,209,200,159,245,193,206,223,255,255,221,234,221,211,220,232,246,0
3,202,228,224,221,211,211,214,205,205,205,220,240,80,150,255,229,221,188,154,191,210,204,209,222,228,225,0
98,233,198,210,222,229,229,234,249,220,194,215,217,241,65,73,106,117,168,219,221,215,217,223,223,224,229,29
75,204,212,204,193,205,211,225,216,185,197,206,198,213,240,195,227,245,239,223,218,212,209,222,220,221,230,67
48,203,183,194,213,197,185,190,194,192,202,214,219,221,220,236,225,216,199,206,186,181,177,172,181,205,206,115
0,122,219,193,179,171,183,196,204,210,213,207,211,210,200,196,194,191,195,191,198,192,176,156,167,177,210,92
0,0,74,189,212,191,175,172,175,181,185,188,189,188,193,198,204,209,210,210,211,188,188,194,192,216,170,0
2,0,0,0,66,200,222,237,239,242,246,243,244,221,220,193,191,179,182,182,181,176,166,168,99,58,0,0
0,0,0,0,0,0,0,40,61,44,72,41,35,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0
```

여기서 0 은 사진의 흰색 픽셀에 해당되며, 값이 255 에 가까워질수록 검은색에 가까운 색을 나타냅니다.

주어진 데이터 파일을 파이썬 코드 파일과 같은 경로에 옮기고, 다음 파이썬 코드로 파일을 열어 string 형태로 변수에 저장할 수 있습니다:

```
img_file = open('MARG Python Assignment 3 (Data).csv', 'r')
img = img_file.read()
img_file.close()
```

이미지 인식을 하는 neural network 를 만들고 있는 상황이라고 가정 해 봅시다. 이러한 28X28 흑백 이미지 파일이 수 천개가 있으며, 이 수 천개의 이미지를 이 neural network 에 입력으로 주어서 훈련을 시켜야 합니다. 그러나 이 neural network 는 아쉽게도 28X28 크기의 csv 파일을 그대로 입력 받지 못한다고 합니다. 따라서 여러분이 할 일은 이러한 28X28 크기의 흑백 이미지 파일을 입력 조건에 맞게 일괄적으로 전처리 해 주는 파이썬 프로그램을 작성하는 것입니다. 인풋 조건은 다음과 같습니다:

1. 196 개(14X14)의 실수로 되어 있음
2. 각 실수는 0.0~1.0 의 크기를 가짐
2. 모든 정보는 각 실수를 쉼표(,)로만 구분한 하나의 .txt 파일에 있어야 함

즉, 여러분이 작성하는 프로그램은 28X28 크기의 흑백 사진을 인풋으로 받아서 14X14 크기로 압축시키고, 0~255 의 정수 스케일을 0.0~1.0 의 실수 스케일로 변환하며, 28 개의 줄에 각각 28 개의 정수 값이 저장되어 있던 기존 파일 형식을 1 개의 줄에 196 개의 실수 값을 저장하는 형식으로 바꾸어 .txt 파일로 출력하는 것입니다. 자유롭게 구현 해 보세요!

참고) 이미지 압축은 여러 알고리즘이 있으나, 이번 과제에서는 간단하게 픽셀을 삭제하는 방식으로 하세요. 즉, 이번 과제는 다음과 같이 하시면 됩니다:

0, 35, 37, 35, 34, 13, 1 → 0, 37, 34, 1

## Question 2. 이미지 데이터 전처리 Challenge (optional)

1 번과 동일한 기능을 수행하는 프로그램을 5 줄의 파이썬 코드만으로 완성해 보세요. 단, 세미콜론(;)을 사용하는 것은 금지이며, 기본적인 파이썬 문법만 사용해야 합니다.