

# Data Science : Assignment 3 Clustering

2018008504 손승일

## 환경 :

VsCode에서 python으로 구현

## 실행방법 :

python clustering.py 'input data' 'N의값' 'Eps값' 'MinPts값'

입력하면 결과값으로 input1.txt의 경우 input1\_cluster\_0 ~ input1\_cluster\_7까지 생성되고 input2의 경우에는 0~4, input3에 대해서는 0~3이 생성된다.

```
PS C:\Users\손승일\Desktop\학교\데사\Assignment3> python clustering.py input1.txt 8 15 22
PS C:\Users\손승일\Desktop\학교\데사\Assignment3> python clustering.py input2.txt 5 2 7
PS C:\Users\손승일\Desktop\학교\데사\Assignment3> python clustering.py input3.txt 4 5 5
```

## 구현 함수:

check\_core(idx) : 해당 idx의 data가 core인지 아닌지 판별하는 함수. Core일 경우에는 directly density reachable list로 만듦

init\_core() : 모든 data에 대해 core임을 판별하고 data마다 directly density reachable list를 만든다 (core일 경우만)

retrieve\_density\_reachable(p) : density\_reachable한 data들을 찾는 함수.

DBscan() : core list를 돌면서 density\_reachable한 모든 data를 한 cluster로 만들고 그걸 cluster\_list에 추가

sort\_cluster() : n개가 넘는 cluster가 존재할 경우 내림차순으로 정렬해서 n개를 끊어야 되서, 가장 큰 cluster부터 내림차순으로 index를 반환하게 하는 함수

read (name) : input.txt로부터 data를 읽는 함수

write (input\_num) input.txt를 읽고 그걸 DBscan을 진행하고 output file을 작성함

## 결과값

```
C:\Users\손승일\Desktop\화학 교재\데사\Assignment3>PA3.exe input1
98.97037점
C:\Users\손승일\Desktop\화학 교재\데사\Assignment3>PA3.exe input2
94.89474점
C:\Users\손승일\Desktop\화학 교재\데사\Assignment3>PA3.exe input3
99.97736점
```

cmd창에서 PA3.exe input1, input2, input3를 각각 입력하면 그 점수가 나온다