2018년 1학기 IT 집중교육 1

Homework 5\_2

\* 제출파일: 학번\_이름\_5-2.ipynb

\* 제출기한: 4/12(목) 11:59p.m.

\* 아주bb에 제출

\* 코드 copy 시 0점

1. **Clustering**
2. KNN으로 만든 matrix(row-user, column-movie, element-rating)를 이용하여 비슷한 성향의 user별 clustering을 수행한다. 이 때 KNN으로 만든 matrix는 knn\_ratings.csv 파일로 제공되며, homework5\_2에서 knn\_ratings.csv파일을 로드하여 사용하면 된다. (matrix의 이름 rating) – homework5\_1과 homework5\_2를 하나의 .ipynb파일에 작업하지 말 것.
   1. movies.csv파일을 이용하여 movie(row)\*genre(column), element: counting genre, matrix를 만든다(matrix의 이름 genre, genre의 정보(18개)는 README 파일 참조).
   2. rating matrix에서 비슷한 성향의 user를 군집화 시키기 위해 Kmeans clustering 알고리즘을 이용한다(k=5로 설정).
   3. 각 cluster 마다 전체 movie rating의 평균을 구한다.
   4. 각 cluster 마다 평균 rating이 높은 20개의 movie를 구하고, genre matrix를 이용하여 각 cluster가 가장 선호하는 genre를 구한다.
3. markdown을 이용하여 5개의 cluster 마다 선호하는 영화의 장르(총 18개-README 파일 참조, 중복되어도 괜찮음)를 정의한다. 그리고 분석 과정, 해당 결과에 대한 해석을 서술한다.

ex)

1번 cluster : Action

2번 cluster : Romance

…

* markdown은 jupyter notebook에서 새로운 cell을 생성한 후 ‘m’을 누르거나 cell의 형식을 Markdown으로 바꾸면 사용할 수 있다.