2018년 1학기 IT 집중교육 1

Homework 1

**추천시스템 과제**

Movielens data는 영화를 관람한 user들이 영화에 대한 평점을 매긴 것에 대한 데이터이다. 이번 추천 시스템 해결 사항은 데이터를 조작하여 추천시스템에 적합한 matrix를 생성하고, 이 생성한 matrix를 바탕으로 KNN을 이용하여 특정 유저에 적합한 영화를 추천해주거나 matrix factorization 기법을 바탕으로 아직 평점을 매기지 못한 영화에 대하여 예측을 한다. 그후 제시된 평가 방법에 따라 성능에 대한 순위를 정해 평가 한다.

\* 코드 copy할 경우 0점 처리

\* 제출기한 04/10 까지 결과물을 아주BB에 제출

실습시간에 주어진 주피터를 바탕으로 작성한다.

**KNN**

1. 주어진 movielens data( ml-latest-small.zip )를 바탕으로 index = 유저, column = 영화, element = 평점인 matrix 생성. 평점을 안 매긴 경우 0으로 채움. ( 올바른 생성시 2점, 아닐 경우 0점)
2. 해당 문서

<http://scikit-learn.org/stable/modules/generated/sklearn.neighbors.NearestNeighbors.html>

를 참고하여 NearestNeighbors 메소드의 파라미터를 적절히 tuning 하여 제시된 검증 방법에 따라서 가장 높은 성능을 내는 파라미터 찾은 후 ipynb 파일에 결과값 출력. 설득을 위한 주석설명과 시각화 기법을 잘 사용하여 해당 파라미터가 합리적인 파라미터인 것을 설명한다. (결과값 출력 성공 시 1점, 시각화와 주석처리 보너스 1점)

제출해야 할 결과물 : 학번이름\_추천시스템\_KNN.ipynb .

**Matrix Factorization**

1. Knn의 (1)에서 만든 matrix data를 바탕으로 과제 진행. 과제시간에 matrix\_factorization 안의 공간을 채워서 newP, newQ가 잘 생성되는지 확인. Step = SGD의 횟수, R = 실제 평점 matrix, P,Q = factorization matrix, alpha =learning late, beta = regularization term. 제시한 return 값은 Q.T 이지만 Q로 바꾸어서 해도 무방. (올바르게 작동 시 3점, 아닐경우 0점 )
2. 파라미터를 잘 조정하여 제시된 검증방법에서 가장 높은 성능을 내도록 구현. (주석처리와 시각화를 통하여 0~2점 부여,)

제출해야 할 결과물 : 학번이름\_추천시스템\_MatrixFactorizaion.ipynb .