

# Project 3: Deep learning based Recommendation

Due: 2021/05/30, 11:59 PM

## 프로젝트 목표

Deep learning based 알고리즘 구현

## Tasks

### ENV

Numpy, Pandas 외 비슷한 종류의 다른 라이브러리를 사용하거나 사용하지 않고 구현해도 무방

- Python 3.7
- Numpy (Optional)
- Pandas (Optional)

Numpy, Pandas 외 비슷한 종류의 다른 라이브러리를 사용하거나 사용하지 않고 구현해도 무방 새로운 라이브러리는 반드시 requirements.txt에 명시 (**버전도 명시**)

**주의사항** : 환경 전체를 freeze를 해서 requirements.txt를 제출이 아닌 반드시 필요한 패키지만 버전을 포함하여 제출 (관련 포맷 [링크](#) 참조)

### Dataset

- Movielens dataset
  - rating\_train.csv: 사용자가 영화에 대해 평가한 평점 (학습용 데이터)
  - Movies\_w\_imgurl.csv: 영화의 장르 등 정보 (학습 보조 데이터)
  - Tags.csv: 사용자의 영화에 대한 태그 정보 (학습 보조 데이터)
  - rating\_val.csv: 사용자가 영화에 대해 평가한 평점 (검증용 데이터)
  - rating\_test.csv: 사용자가 영화에 대해 평가한 평점 (채점용 데이터, 제공되지 않음)
  - Tags.csv, Movies\_w\_imgurl.csv 데이터의 경우 필요에 따라 자유롭게 사용하면 됩니다. Test시에는 사용 불가.

## Task

1. 딥러닝 기반 알고리즘 선정
2. 알고리즘 구현 및 학습
  - A. 데이터 전처리
  - B. 모델 학습
3. 하이퍼파라미터 튜닝

## 평가기준

1. 특정 user\_id, movie\_id에 대한 prediction\_score와 실제 rating 사이의 RMSE
2. 포맷 준수 여부
  - A. input.txt의 user\_id, movie\_id에 해당하는 점수출력
  - B. output.txt의 필드(user\_id, movie\_id, prediction\_score)간 구분자는 공백 없이 콤마(,)를 사용
  - C. prediction\_score의 경우, 소수점 8번째까지 반올림
  - D. output.txt 파일은 main.py 파일과 같은 위치에 출력
3. 알고리즘 구동 시간
  - A. 입력 1405개에 대해서 2시간 내 output.txt 출력 필수
  - B. 테스트환경
    - i. CPU: 8 Core
    - ii. RAM: 32GB
    - iii. GPU: Geforce GTX 1080
    - iv. CUDA 11
4. Requirements.txt 준수 여부
  - A. 이번 과제 특성상 requirements.txt가 올바르지 않아서 생기는 문제에 대해서는 패널티가 있습니다.
  - B. 반드시 [링크](#)를 확인하시고 'pip install -r requirements.txt'로 라이브러리를 설치 가능하도록 명기 바랍니다.
  - C. 또한 freeze로 넣은 requirements.txt로 불필요한 라이브러리 명기는 모두 지워 주시기 바랍니다.

## example

```
# input.txt format
# 단순 한줄당 user_id, movie_id
# 해당 example은 정답과 무관한 순수 입출력 형식 예시입니다

1,31
2,10
3,1235
```

```
4,10
```

```
...
```

```
# output.txt format # 순서대로 (user_id, movie_id, prediction_score) 출력
```

```
1,31,2.50000000
```

```
2,10,4.00000000
```

```
3,1235,4.00000000
```

```
4,10,4.00000000
```

```
...
```

## 제출

### 제출방법

- ETL 과제에 제출
- 본 과제 제출물 특성상 메일이 아닌 ETL로 제출합니다.
- Source code와 Report를 함께 zip파일로 묶어 첨부하여 제출합니다. (e.g. PRJ3\_202100000\_홍길동.zip)
- 1 day delay 10% 감점, 2 days부터 0점

### Source Code

- main.py 파일은 다음과 같은 명령어를 입력했을 때, input.txt에 해당하는 output.txt 파일을 출력해야 합니다.
- 소스코드 내에는 학습 코드도 포함해야 합니다.
- 학습 후 계산된 파라미터는 'param.data'로 저장 후 함께 제출해야 합니다.
- 최종 제출본에서는 main.py 실행 시 학습된 파라미터를 읽어 수행되어야 합니다.
- 제공한 input.txt (1405개)기준 2시간이내에 완료해야 합니다.

```
# main.py와 input.txt가 같은 위치에 존재할 때  
python main.py
```

### Report

- 다음과 같은 내용을 보고서에 간단하게 작성해 주세요.
  - 구현한 딥러닝 기반 알고리즘 설명

■ 학습과정 설명

◆ 전처리 과정 분석

◆ 학습 (Loss 그래프 등) 분석

◆ 하이퍼파라미터 튜닝에 따른 성능 개선(성능 표 등) 분석

■ 본 프로젝트에 대한 피드백 의견

● 길이는 최대 A4 4장까지이며 작성 후 PDF형식으로 제출을 원칙으로 합니다.

● 파일 제목은 다음과 같은 양식을 준수해야합니다.

(e.g. PRJ3\_202100000\_홍길동.pdf)

# 제출 상태(구현 파일에 대한 이름 및 파일 개수 제한 없음)

PRJ3\_202100000\_홍길동.zip

├── PRJ3\_202100000\_홍길동.pdf

├── data

│ ├── rating\_train.csv

│ ├── Movies\_w\_imgurl.csv

│ ├── Tags.csv

│ └── rating\_val.csv

├── main.py

├── param.data

└── requirements.txt