IPA 주관 인공지능센터 기본(fundamental) 과정

- **GitHub link: here**
- **E-Mail: windkyle7@gmail.com**

In [1]:

```
import pandas as pd
from surprise import Dataset

path = '../../workspace/data/ml-100k'
cols_users = ['user_id', 'age', 'sex', 'occupation', 'zip_code']
cols_ratings = ['user_id', 'movie_id', 'rating', 'unix_timestamp']
cols_movies = [
    'movie_id', 'title', 'release_date', 'video_release_date', 'imdb_url'
]

users = pd.read_csv(path + '/u.user', sep='|', names=cols_users)
ratings = pd.read_csv(path + '/u.data', sep='\t', names=cols_ratings)
movies = pd.read_csv(path + '/u.item',
                     sep='|',
                     usecols=range(5),
                     encoding="latin1",
                     names=cols_movies)
lens = pd.concat([movies, ratings, users], axis=1)
lens.sample(5)
```

Out[1]:

	movie_id	title	release_date	video_release_date	imdb_url	user_id	movie_id	rating	unix_time
72164	NaN	NaN	NaN	NaN	NaN	504	216	4	8878
39608	NaN	NaN	NaN	NaN	NaN	445	1010	1	8912
9780	NaN	NaN	NaN	NaN	NaN	374	581	4	8809
30100	NaN	NaN	NaN	NaN	NaN	161	50	2	8911
93034	NaN	NaN	NaN	NaN	NaN	899	177	3	8841

In [2]:

```
lens.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100000 entries, 0 to 99999
Data columns (total 14 columns):
movie_id              1682 non-null float64
title                 1682 non-null object
release_date          1681 non-null object
video_release_date    0 non-null float64
imdb_url              1679 non-null object
user_id               100000 non-null int64
movie_id              100000 non-null int64
rating                100000 non-null int64
```

```
unix_timestamp         100000 non-null int64
user_id                943 non-null float64
age                    943 non-null float64
sex                    943 non-null object
occupation             943 non-null object
zip_code               943 non-null object
dtypes: float64(4), int64(4), object(6)
memory usage: 10.7+ MB
```

In [3]:

```
from surprise import Reader

reader = Reader(rating_scale=(1, 5))
ratings_dict = {
    'itemID': [1, 1, 1, 2, 2],
    'userID': [9, 32, 2, 45, 'user_foo'],
    'rating': [3, 2, 4, 3, 1]
}
df = pd.DataFrame(ratings_dict)
data = Dataset.load_from_df(df[['userID', 'itemID', 'rating']], reader)
data
```

Out[3]:

```
<surprise.dataset.DatasetAutoFolds at 0x7f24b40a0b70>
```

In [4]:

```
data.df
```

Out[4]:

	userID	itemID	rating
0	9	1	3
1	32	1	2
2	2	1	4
3	45	2	3
4	user_foo	2	1

In [5]:

```
lens.columns
```

Out[5]:

```
Index(['movie_id', 'title', 'release_date', 'video_release_date', 'imdb_url
',
       'user_id', 'movie_id', 'rating', 'unix_timestamp', 'user_id', 'age',
       'sex', 'occupation', 'zip_code'],
      dtype='object')
```

In [6]:

```
lens[['movie_id', 'user_id', 'rating']][:5]
```

Out[6]:

	movie_id	movie_id	user_id	user_id	rating
0	1.0	242	196	1.0	3
1	2.0	302	186	2.0	3
2	3.0	377	22	3.0	1
3	4.0	51	244	4.0	2
4	5.0	346	166	5.0	1

In [7]:

```
lens = pd.merge(pd.merge(movies, ratings), users)
lens.sample(5)
```

Out[7]:

	movie_id	title	release_date	video_release_date	imdb_url	user_id	ratin
9261	173	Princess Bride, The (1987)	01-Jan-1987	NaN	http://us.imdb.com/M/title-exact?Princess%20Br...	41	
77535	519	Treasure of the Sierra Madre, The (1948)	01-Jan-1948	NaN	http://us.imdb.com/M/title-exact?Treasure%20of...	474	
48682	290	Fierce Creatures (1997)	10-Jan-1997	NaN	http://us.imdb.com/M/title-exact?Fierce%20Crea...	733	
36952	5	Copycat (1995)	01-Jan-1995	NaN	http://us.imdb.com/M/title-exact?Copycat%20(1995)	378	
23991	11	Seven (Se7en) (1995)	01-Jan-1995	NaN	http://us.imdb.com/M/title-exact?Se7en%20(1995)	99	

In [8]:

```
lens[['movie_id', 'user_id', 'rating']][:5]
```

Out[8]:

	movie_id	user_id	rating
0	1	308	4
1	4	308	5
2	5	308	4
3	7	308	4
4	8	308	5

In [9]:

```
from surprise import SVD
```

```
from surprise import SVD
import inspect
# inspect.getsource(SVD)
inspect.getmodule(SVD)
```

Out[9]:

```
<module 'surprise.prediction_algorithms.matrix_factorization' from '/home/user/workspace/.venv/lib/python3.6/site-packages/surprise/prediction_algorithms/matrix_factorization.cpython-36m-x86_64-linux-gnu.so'>
```