Code Documentation - Baselines

The baseline experiments performed are a random recommender, popularity based recommender, and item-item recommender. The code for the random recommender, popularity based recommender and their subsequent comparison with a collaborative filtering method and evaluation are borrowed from Claire Longo (statisticianinstilettos)'s recmetrics libary's examples. GitHub link: https://github.com/statisticianinstilettos/recmetrics

File: Baseline_Recommenders

Dependencies and their installation:

Recmetrics (a library of evaluation metrics for recommender systems) - pip install recmetrics
Surpriselib (a libarry that helps create recommender systems easily) - pip install surprise
Numpy - pip install numpy
Pandas - pip install pandas

Differences in Implementation:

The data is downloaded from the Grouplens website and ratings are extracted to perform experiments. The actual code uses the Movielens 27M dataset but our implementation uses the 100k dataset.
Additionally, the ratings are not cut off at 1 star - in our experiments we use all ratings. The random recommender does random sampling and generates 10 random movie IDs for each user.
Catalogue coverage and recall are calculated using the recmetrics library functions.
The actual function to print the names of the movies (and just not tabulate the recommended movie IDs) using a dictionary was absent in the original implementation.

We observe qualitative results in the form of top-N recommendations for a selected user, user 5. Some of the movies that this user has watched are: ['Toy Story (1995)',
 'GoldenEye (1995)',
 'From Dusk Till Dawn (1996)',
 'Muppet Treasure Island (1996)',
 'Rumble in the Bronx (1995)',
 'Birdcage, The (1996)',
 'Batman Forever (1995)',
 'Star Wars (1977)',
 'Stargate (1994)',
 'Santa Clause, The (1994)',
 'While You Were Sleeping (1995)',

'Forrest Gump (1994)',
'Four Weddings and a Funeral (1994)',
'Fugitive, The (1993)',
'Blade Runner (1982)',
'Home Alone (1990)',
'Aladdin (1992)',
'Silence of the Lambs, The (1991)',
'Snow White and the Seven Dwarfs (1937)',
'Fargo (1996)',
'Mystery Science Theater 3000: The Movie (1996)',
'Operation Dumbo Drop (1995)',
'Independence Day (ID4) (1996)',
'2001: A Space Odyssey (1968)',
'Love Bug, The (1969)',
'Die Hard (1988)',
'Lawnmower Man, The (1992)',
'Willy Wonka and the Chocolate Factory (1971)',
'Fish Called Wanda, A (1988)',
"Monty Python's Life of Brian (1979)",
'On Golden Pond (1981)',
'Return of the Pink Panther, The (1974)',
'Private Benjamin (1980)',
'Monty Python and the Holy Grail (1974)',
'Empire Strikes Back, The (1980)',
'Princess Bride, The (1987)',
'Raiders of the Lost Ark (1981)',
'Aliens (1986)',
'Return of the Jedi (1983)',
'Alien (1979)',
'Psycho (1960)',
'Shining, The (1980)',
'Back to the Future (1985)',
'Indiana Jones and the Last Crusade (1989)',
'Pink Floyd - The Wall (1982)',
'When Harry Met Sally... (1989)',
'Nightmare on Elm Street, A (1984)',
'Star Trek: First Contact (1996)',
'101 Dalmatians (1996)',
'Die Hard 2 (1990)',
'Star Trek VI: The Undiscovered Country (1991)',
'Star Trek: The Wrath of Khan (1982)',
'Star Trek III: The Search for Spock (1984)',
'Star Trek IV: The Voyage Home (1986)',
'Batman Returns (1992)',

'Jaws (1975)',
'Men in Black (1997)',
'George of the Jungle (1997)',
'Star Trek: Generations (1994)',
'Last Action Hero (1993)',
'Man Without a Face, The (1993)',
'Mrs. Doubtfire (1993)',
'Serial Mom (1994)',
'Striking Distance (1993)',
'Super Mario Bros. (1993)',
'Three Musketeers, The (1993)',
'Little Rascals, The (1994)',
'Brady Bunch Movie, The (1995)',
'Ghost (1990)',
'Batman (1989)',
'Pinocchio (1940)',
'Mission: Impossible (1996)',
'Spy Hard (1996)',
'Close Shave, A (1995)',
'Nutty Professor, The (1996)',
'Parent Trap, The (1961)',
'Cinderella (1950)',
'Mary Poppins (1964)',
'Alice in Wonderland (1951)',
"William Shakespeare's Romeo and Juliet (1996)",
'Aladdin and the King of Thieves (1996)',
'E.T. the Extra-Terrestrial (1982)',
'Children of the Corn: The Gathering (1996)',
'Bob Roberts (1992)',
'Transformers: The Movie, The (1986)',
'To Kill a Mockingbird (1962)',
'American Werewolf in London, An (1981)',
"Amityville 1992: It's About Time (1992)",
'Amityville 3-D (1983)',
'Amityville: A New Generation (1993)',
'Amityville II: The Possession (1982)',
'Amityville Horror, The (1979)',
'Amityville Curse, The (1990)',
'Birds, The (1963)',
'Blob, The (1958)',
'Body Snatcher, The (1945)',
'Burnt Offerings (1976)',
'Omen, The (1976)',
'Star Trek: The Motion Picture (1979)',

'Star Trek V: The Final Frontier (1989)',
'Grease (1978)',
'Jaws 2 (1978)',
'Jaws 3-D (1983)',
'Bastard Out of Carolina (1996)',
"Jackie Chan's First Strike (1996)"']

This list appears to be a combination of children's movies, action movies, sci-fi, and horror.

For random recommender, the recommended movies for user 5 turn out to be:
GoldenEye (1995)
On Golden Pond (1981)
Beautiful Girls (1996)
American Werewolf in London, An (1981)
Magic Hour, The (1998)
Priest (1994)
Paris, Texas (1984)
With Honors (1994)
Thin Man, The (1934)
The Innocent (1994)

Popularity-based recommender takes the movies with the most ratings and recommends the same 10 movies for every user. As expected, catalogue coverage is very poor.
For popularity-based recommender, the recommended movies for user 5 turns out to be:
Legends of the Fall (1994)
George of the Jungle (1997)
Heavy Metal (1981)
GoodFellas (1990)
Breakdown (1997)
Marvin's Room (1996)
Evita (1996)
GoldenEye (1995)
In & Out (1997)
Cable Guy, The (1996)

Files: Item-Item_Recommender_Normalization.ipynb and Item-Item_Recommender_Without_Normalization.ipynb

Dependencies and how to install them:

scikit learn - pip install scikit-learn
Numpy - pip install numpy
Pandas - pip install pandas
scipy - pip install scipy

Differences in implementation:

The code builds on the majority borrowed from the beginner-friendly implementation by Rohith Kumar Poshala (https://github.com/rposhala/Recommender-System-on-MovieLens-dataset). However, the Bayesian normalization and its implementation in coding was inspired by Jill Cates' implementation of item-item recommenders in this tutorial - https://github.com/topspinj/recommender-tutorial/blob/master/part-1-item-item-recommender.ipynb

Item-Item recommenders are a way to model the "Movies like X" or "Because you watched Y" which works based on a distance metric such as cosine or euclidean. The scikit library's unsupervised knn method has been used.
sklearn.neighbours.NearestNeighbours - is used to compute the cosine distances between movies by passing a sparse matrix, ie, our user-item matrix which contains user ratings for all movies.
Given any input movie, it computes the nearest neighbours of that movie in the 1682-dimensional space.

For the imput movie "Toy Story (1995)", the movies recommended (using raw ratings) are:
Star Wars (1977)
Return of the Jedi (1983)
Independence Day (ID4) (1996)
Rock, The (1996)
Mission: Impossible (1996)
Willy Wonka and the Chocolate Factory (1971)
Star Trek: First Contact (1996)
Fargo (1996)
Jerry Maguire (1996)

Raiders of the Lost Ark (1981)

For the imput movie "Toy Story (1995)", the movies recommended (using Bayesian-normalised ratings) are:

Star Wars (1977)
Independence Day (ID4) (1996)
Return of the Jedi (1983)
Rock, The (1996)
Fargo (1996)
Willy Wonka and the Chocolate Factory (1971)
Mission: Impossible (1996)
Twelve Monkeys (1995)
Star Trek: First Contact (1996)
Jerry Maguire (1996)

The order of recommended movies is rearranged and one movie is replaced after normalization of ratings.