

ts2956 Tianyi Shao Homework 4

April 15, 2016

Problem 1

1. The code for calculate K-means objective function is **Kmeans.m**, which needs an input of K, and outputs the randomed Data, their clusters and plot the objective function values. The graphs are **K=2,3,4,5.jpg** and they all look nice!
2. b. Use **plt.m** to plot the clusters when K=3 and K=5. The graphs are **K=3 Clusters.jpg** and **K=5 Clusters.jpg**, data points from different clusters are labeled with different colors.

Problem 2

The code for constructing M and the index set Ω is **constructM.m**, which needs to input the given movie_ratings.mat, outputs M and Ω . The matrix factorization process is **map.m**, inputting M, Ω and the test set, outputting the result U,V and error.

1. Plot the error data and get the graph **RMSE.jpg**.
2. The three movies I pick is **Golden Eye(2)**, **Apollo 13(28)**, **Forrest Gump(69)**. Use **nearest.m** to find the top 5 nearest films to these three films. 5 films similar to **Golden Eye** are: **The Ghost and The Darkness**, **How to Make an American Quilt**, **Outbreak**, **Matilda**, **Under Siege**. 5 films similar to **Apollo 13** are: **The Rainmaker**, **Indiana Jones and the Last Crusade**, **The Terminator**, **Forrest Gump**, **Rosewood**. 5 films similar to **Forrest Gump** are: **The Rainmaker**, **The Hunt for Red October**, **The Game**, **Apollo 13**, **Mr.Holland's Opus**.
3. Using code **number.m**, we get the top 5 centroids are:

```
[-0.1096 -0.1096 -0.1096 -0.1096 -0.1096 -0.1096 -0.1096 -0.1096 -0.1096 -0.1096],  
[-0.1021 -0.1021 -0.1021 -0.1021 -0.1021 -0.1021 -0.1021 -0.1021 -0.1021 -0.1021],  
[0.0261 0.0261 0.0261 0.0261 0.0261 0.0261 0.0261 0.0261 0.0261 0.0261],  
[0.0027 0.0027 0.0027 0.0027 0.0027 0.0027 0.0027 0.0027 0.0027 0.0027],  
[-0.0338 -0.0338 -0.0338 -0.0338 -0.0338 -0.0338 -0.0338 -0.0338 -0.0338 -0.0338].
```

And the number of their members are: 75,88,89,96,126.

Ten movies to centroids 1,2 and 5 are the same, they are: **The Bloody Child**, **Mat'i syn**, **Butterfly Kiss**, **Getting Away with Murder**, **August**, **Coldblooded**, **Death in the Garden**, **He Walked by Night**, **Girls Town**, **You So Crazy**.

Ten movies to centroids 3 and 4 are the same, they are: **The Eighth Day**, **Dadetown**, **Hana-bi**, **The Big One**, **Angel Baby**, **Some Mother's Son**, **Brothers in Trouble**, **The Butcher Boy**, **The Butcher Boy**, **The Spanish Prisoner**.

4. sing code **numberV.m**, similarly, we get the top 5 centroids in films: -0.0459I,-0.0415I, 0.0014I, -0.0075I, -0.0046I. And the number of their members are: 149,158,184,186,223.

Ten movies to all the 5 centroids are the same: **L.A. Confidential**, **Four Days in September**, **Pather Panchali**, **Someone Else's America**, **A Close Shave**, **Good Will Hunting**, **A Great Day in Harlem**, **Saint of Fort Washington**, **Star Kid**, **Entertaining Angels: The Dorothy Day Story**.