EE-219 Project 3

Ashish Shah(UID)

Ayush Dattagupta(UID)

Shrey Agarwal (004943082)

Varun Saboo (505028591)

February 22, 2018

Project 3: Collaborative Filtering

**Objective –**

In this project, we will build a recommendation system to predict the ratings of the movies in the MovieLens dataset.

**Dataset –**

We work with MovieLens dataset which is a collection of movies and users along with their ratings and timestamp. Each movie belongs to certain genres which we are not considering in this project.

Question 1:

**Compute Sparsity:**

Sparsity is: 0.9835608583913366

Question 2:

**Plot the histogram showing frequency of the rating values:**

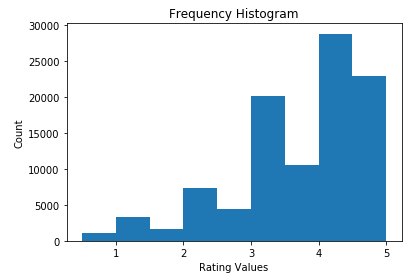


Figure 1: Histogram of frequency

From the histogram, we can infer that majority of the movies have got higher ratings.

Question 3:

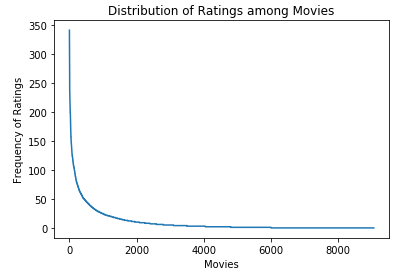
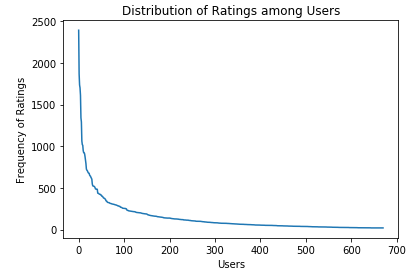
**Plot the distribution of ratings among movies:**

Figure 2: Distribution of ratings versus movies

As we can observe, the number of movies that receive ratings decrease exponentially. That is, there are very few movies that have high number of ratings.

Question 4:

**Plot distribution of ratings among users:**



Question 5:

**Explain salient features of distribution in Question 3:**

The number of movies are: 9066.

Out of which, only 3099 movies have >5 ratings in total.

Total of 3063 movies have only one rating.

The above statistics along with the Sparsity of 0.98 tells us that the data is very skewed and hence the need for a method to predict ratings arises.

Question 6:

**Compute Variance of rating values received by each movie:**

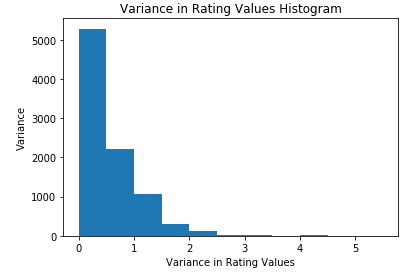
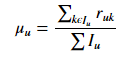


Figure 3: Variance in Rating Values for each movie

The above histogram shows us that the ratings received by all users for each movie is in between [0, 1] majorly and there are very few movies with high variance (> 1). The lesser the variance in ratings, the precise the prediction of ratings will be.

Question 7:

**Write formula for µu in terms of Iu and ruk**



Question 8:

**In plain words, explain the meaning of Iu ∩ Iv. Can Iu ∩ Iv = Ø**

Iu ∩ Iv: This means we are trying to find the set of items that both the users u and v have rated.

Can Iu ∩ Iv = Ø: Yes, it can!

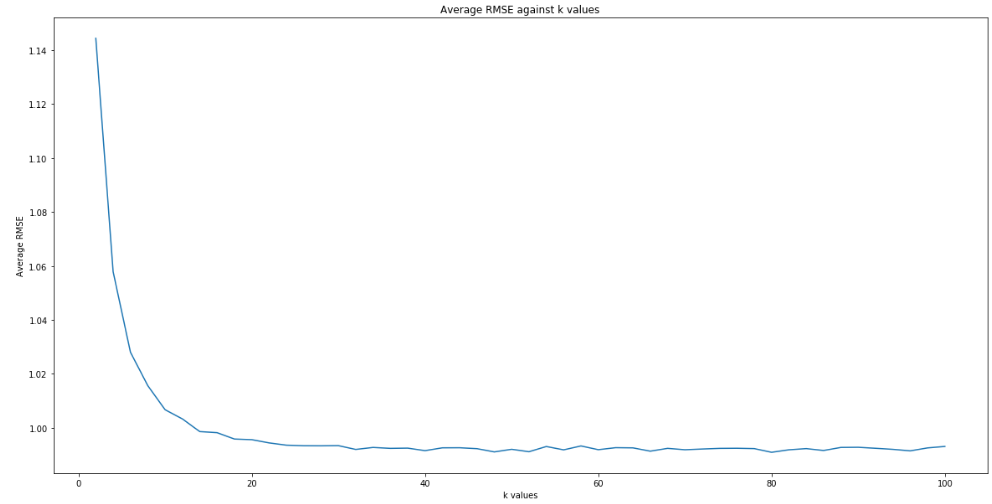
Iu ∩ Iv = Ø: If the set of items that both the users have rated is empty, it means that the users u and v have no items common to each other. This also indicates that the users u and v are very unlikely to be neighbours.

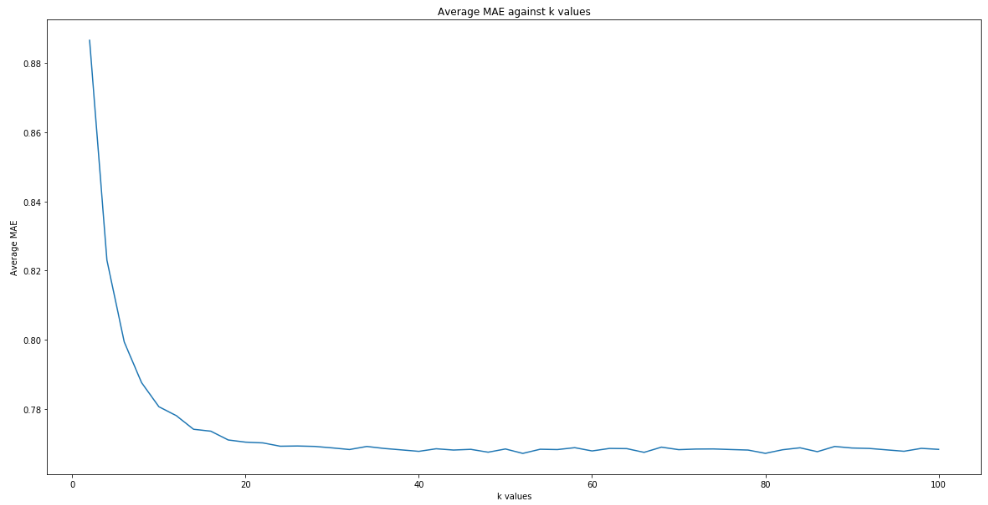
Question 9:

**Explain:**

Question 10:

**Design a k-NN collaborative filter to predict the ratings of the movies in the MovieLens dataset and evaluate its performance using 10-fold cross validation. Plot average RMSE (Y-axis) against k (X-axis) and average MAE (Y-axis) against k (X-axis).**





We can observe that the Average MAE and RMSE are converging to a steady state after certain k value which we will determine in the next section. Steady state is reached when the average MAE and RMSE don’t change significantly with increase in k value.

Question 11:

**Find minimum k:**

Minimum k with RMSE= 48 and with MAE = 52. Hence, we can say that minimum k value is around 50. Hence, we use **minimum k=48** for further questions.

Average RMSE for minimum k (48) is: 0.9920527987540382

Average MAE for minimum k (48) is: 0.7684647079132884

Question 12:

**Q**