Assignment 6

Ellen Hsieh

1. Netflix Prize and Bell, Koren, and Volinsky (2010)

(a) Netflix Prize is a competition held by Netflix in order to improve its movie recommendation system, Cinematch^{SM}. Therefore, all the submissions would be judged based on whether they could improve the Netflix's Cinematch^{SM} by 10% or more. The criterion function for the judgment is root-mean-square error (RMSE). The smaller the RMSE, the better the improvement. After browsing the Netflix Prize website (https://www.netflixprize.com/rules.html), I cannot find any specific cutoffs for this competition since Netflix would review all the submissions written in English.

(b) The most commonly used method for predicting ratings on movies is nearest neighbors, which predict rating for item by a user using a weighted average rating of similar items by the same user. In this article, the authors (Bell, Robert M., Yehuda Koren, and Chris Volinsky, 2010, p.25) mention that the similarity in this method is usually measured via Pearson correlation, cosine similarity, or other metric calculated on the ratings.

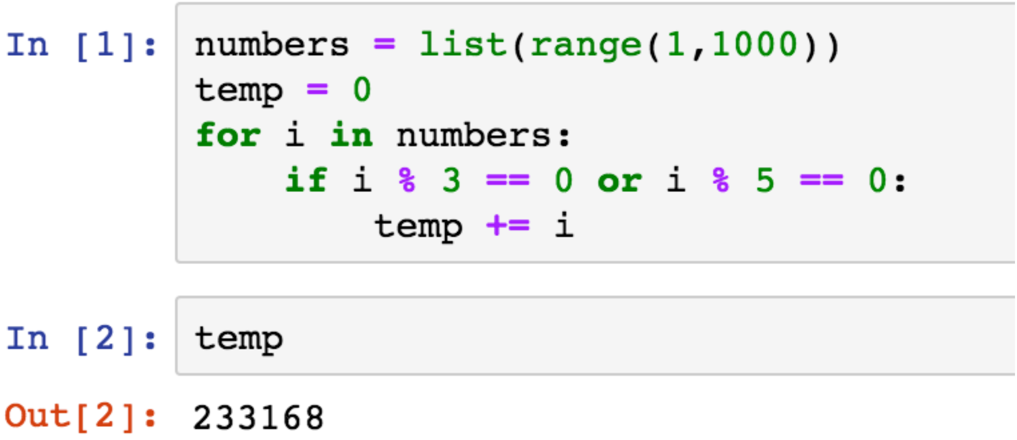
(c) One of the predictive model in the best predictive models in the Netflix Prize is matrix factorization. The characteristic of matrix factorization is that it uses a smaller number of d-dimensional vectors of latent factors to predict a larger number of items and users, in this case, the items are movies and the users are subscribers of Netflix (Bell, Robert M., Yehuda Koren, and Chris Volinsky, 2010, p.26). For movies, a factor might reflect its characteristics such as the amount of violence, whether it is irony or satire, drama or comedy. In the same dimensional space, there is a vector which would reflect the user's taste for items that has higher score on the corresponding factor.

2. Collaborative problem solving: Project Euler

(a) My Project Euler username and friend key are as following:

ellenhsieh: 1409794_YOZz4jyn8VmzQrQPDXAjeavCDEalYLFf

(b) The problem I chose to solve is problem 1 (Multiples of 3 and 5). The following are my code and answer to the problem:



```
In [1]:  numbers = list(range(1,1000))
         temp = 0
         for i in numbers:
             if i % 3 == 0 or i % 5 == 0:
                 temp += i

In [2]:  temp

Out[2]:  233168
```

(c) Three awards that attract me the most are:

(1) The Journey Begins: progress to level 1 by solving twenty-five problems

(2) High Five: Solve the five most recent problems

(3) Gold Medal: The first to solve the problem

Since I am still new to coding, “The Journey Begins” inspires me to challenge myself through solving twenty-five problems to get to a higher level, which would make me feel that I’m making some progress in programming. For “High Five”, solving recent problems just sounds interesting to me since those questions might be pretty new and different from the previous ones. Getting the “Gold Medal” award seems really fascinating since I am a person who loves to challenge myself. Besides, being the first person to solve the problem would be really cool and can make me feel accomplished.

3. Human computation projects on Amazon Mechanical Turk

(a) One of the human computation projects on MTurk is called “Find info from an email v1” which is requested by WatchFlower Systems. The task for the participants is to find the name, provider type, state and country, given an email and website.

(b) The reward for participating in this project is $0.03.

(c) The qualifications of the project are as following:

(1) Watchflower Block has not been granted

(2) HIT approval rate (%) is not less than 85

(3) Watchflower Pending Review has not been granted

(4) Total approved HITs is greater than 1000

(d) The allotted time for this task is 60 minutes. Therefore, in an hour I can only finish one task. The hourly rate of the task is $0.03/hour.

(e) The expiration date of the job is 11/16/2019.

(f) If there are 1 million people participating in the task, the HIT creator, WatchFlower System, would need to spend $30,000 on this project. ($0.03/person * 1,000,000 people = $30,000)

4. Kaggle open calls

On Kaggle, there is an interesting competition called “Quick, Draw! Doodle Recognition Challenge”. The purpose of this competition is to build a better classifier for “Quick, Draw!” dataset so that the drawings in the dataset could be better recognized. The competition sponsor is Google LLC, an American multinational technology company specialized in Internet-related services and products. In this competition the submission would be evaluated by Mean Average Precision @3 (MAP@3). The formula of the function is as following:

$$MAP\ @3 = \frac{1}{U}\sum_{u=1}^{U}\sum_{k=1}^{\min(n,3)} P(K)$$

where U is the number of scored drawings in the test data, P(K) is the precision at cutoff k, and n is the number predictions per drawing. (https://www.kaggle.com/c/quickdraw-doodle-recognition#evaluation)

The total prize for this competition is $25,000. For the first place, the winner team can get $12,000. For the second place and the third places, the winning teams can win $8,000 and $5,000, respectively. The entry deadline and the team merger deadline for this competition is November 27, 2018. The final deadline for the submission is December 4, 2018. The contestants can make their submission through Kaggle Kernels, which would allow them to collaborate with their teammates on it. The restriction on submission file is that it should contain a header and list the key_id (in the test set) with three corresponding predicted words. The predicted words should be separated by the space. However, if the word contains more then one word, the label prediction should replace the space with underscore.

In addition to the rules for the submission format, there are also some requirements for submission code itself. First, no private code sharing is allowed, especially between separate teams, unless a team merger happened. Second, public code sharing is permitted if such action does not violate the intellectual property rights of any third party. However, if the participants choose to publicly share their competition code, then they are required to share it on Kaggle.com to benefit other competitors. Also, the participant who shares the code need to license the shared code under an Open Source Initiative-approved license (www.opensource.org). Finally, using open source code is allowed. Nevertheless, the competitors can only use the open source code that are licensed under Open Source Initiative-approved license.

After the competition is over, its sponsor, Google LLC, might apply the winner's solution to build a better handwriting recognition and improve other related applications such as OCR (Optical Character Recognition), ASR (Automatic Speech Recognition) and NLP (Natural Language Processing) in its own services and products such as Google's phones, tablets, laptops, and its connected home series like Google Home. For instance, Google can improve the performance of its Pixel phones by enhancing its handwriting recognition so that its users can take notes by writing them down instead of typing but still can easily convert their handwriting into text on the device. Besides, the refined OCR or ASR system can strengthen the ability of Google Home to recognize the users' demand and respond to them more accurately.

References

Bell, Robert M., Yehuda Koren, and Chris Volinsky, "All Together Now: A Perspective on the Netflix Prize," Chance, 2010, 23 (1), 24–29.