Assignment #6

MACS 30000, Dr. Evans

Dongcheng Yang

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

(a) The submissions to the Netflix Prize open call contest would be judged compared to Netflix's website movie recommendation system, Cinematch. The criterion is the root mean squared error (RMSE) (Bell et al., 2010, p24) and the function is as follows:

$$\sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}}$$

where $\hat{y}_i$ is the predicted stars and $y_i$ is the actual rating in the test dataset.

Only the teams which managed to improve Cinematch's RMSE by 10% or more could possibly win the prize.

(b) At the beginning of the Netflix Prize contest, Nearest Neighbors was the most commonly used methods for predicting ratings on movies. The predicted star for a movie by a particular user, under the nearest neighbors metric, is a weighted average rating of similar items by the same user (Bell et al., 2010, p25). The advantage of using this method is that the "resulting recommendations are easy to explain". However, the usual caveat also applies since the "similarity metric is arbitrary" (Bell et al., 2010, p26). Besides, for the movies without close neighbors, the predictions might not be convincing (Bell et al., 2010, p26).

(c) One model's low correlation with other models in the hybrids serves as a way to improve the overall prediction (Bell et al., 2010, p28).

## Reference

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

2. Collaborative problem solving: Project Euler

(a) dongcheng: 1408103_rrXKnzpnGTxUQJAEOEW8dsVvyt6Ok66c

(b) The problem that I chose:

problem1: Find the sum of all the multiples of 3 or 5 below 1000.[1]

python code that I used to solve the problem:

```
sum = 0
for i in range(1000):
    if i%3==0 or i%5==0:
        sum+=i
print(sum)
```

Answer: 233168

(c) CC for Continued Commitment: This award is for the programmers who have solved the first two-hundred problems, which is a good indicator of one's dedication and passion on programming.

Big Game Hunter: This award is a good measurement of the programmers' ability to handle hard problems.

Perfection: To get this award, we need to solve every problem, which seems to be an ultimate target of joining this programming game.

3. Human computation projects on Amazon Mechanical Turk

(a) Title: Use Computer Browser(Chrome, FireFox, Safari, EDGE) click the URL[2]

(b) The full payment structure: $0.30 for completing this task

(c) Qualifications Required: Masters has been granted

(d) The allotted time for this task is 60 min. In an hour, I could possibly handle 5 items. Thus, the implied hourly rate is $1.50.

(e) This job will expire on November 16.

(f) If each person is only permitted to complete one single task, then $300,000 is the most this project would cost the HIT creator if 1 million people participated.

---

[1] See the whole problem on https://projecteuler.net/problem=1

[2] See more detains of this project on: https://worker.mturk.com/projects/3YATNUOYWN0NKA8OMESY5UR48Y35RT/tasks?ref=w_pl_prvw

4. Kaggle open calls

(a) username: dongchengecon

(b) The title of the competition is "Two Sigma: Using News to Predict Stock Movements" and it is sponsored by Two Sigma. Two Sigma is not like the typical investment manager. The principles of technology and innovation serve as the basic guideline of this company. Technology and data science have been applied by Two Sigma to financial forecasts for over 17 years.[3] As for the evaluation of the submissions, the overall score of a project could be calculated as the return(mean) divided by risk(standard deviation).[4] The team which wins the first place will be awarded $25,000. $20,000 is for the second place and $15,000 for the third place. Teams taking the fourth through seventh place will receive $10,000 each. Besides the behaviors mentioned in the official competition rules, in this competition, any attempt to take advantage of the information that is outside of the provided dataset, or an attempt to abuse the provided information could be viewed as "cheating".[5] The start date of the competition was September 25 and the deadline for submission is January 8, 2019. Teams are required to select two best submissions themselves and "no auto-selection is provided". The final end date will be July 15, 2019. All submissions will occur through the Kernels environment. The submission file for this competition will span both the historical time period and the future time period for simplicity, which means there is only one “valid” submission file at a given time. Site-wide Kernels limits should be followed during the competition, although the limits change during the whole process. The current limits are: 16 GB Memory, 6 Hours total runtime and 4 CPU cores.

(c) The winning submission answer could be taken advantage of by the sponsoring entity as a reference to improve its existing models for predicting stock prices. It is a brilliant idea to make use of news data to advance the state of research in financial market pricing. If harnessed properly, this wining code could lead us to a new era of financial prediction world.

[3] See more detail of Two Sigma on the website: https://www.twosigma.com

[4] The website link of calculating the final score: https://www.kaggle.com/c/two-sigma-financial-news#evaluation

[5] See the whole description of honor code on the website: https://www.kaggle.com/c/two-sigma-financial-news#Honor-Code