Assignment 6

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

(a)

The aim of Netflix Prize was to find the model that achieved "the greatest improvement in root mean squared error (RMSE) over Netflix's internal algorithm, Cinematch"(Bell et al. (2010): p.24). So only the submissions that had a RMSE 10% lower than the Cinematch level (0.9514) would be judged.

(b)

At the beginning of the Netflix Prize contest, the most commonly used method was nearest neighbors. (Bell et al. (2010): p.25)

(c)

If the correlation between the two models is not high, combining the two models can improve the RMSE. (Bell et al. (2010): p.28)

2. Collaborative problem solving: Project Euler

(a)

Shuyan: 1407389_i3Bp78gEKZ4tDBUwQwhIla9CcEadW9aI

(b)

The problem: Multiples of 3 and 5(https://projecteuler.net/problem=1)

My code:

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

My answer: 233168

(c)

1. Lucky Luke(Solve fifty lucky numbered problems). I like this reward because lucky number problems are interesting and I like to figure out why these numbers are "lucky".
2. Fibonacci Fever(Solve the first twelve Fibonacci numbered problems). I like this reward because I can practise recursive programming through Fibonacci numbered problems.

3. One In A Hundred(Be among the first hundred to solve a problem). I like this reward because it will be an exciting achievement if I can solve a new problem before most people.

3. Human computation projects on Amazon Mechanical Turk

(a)

I selected "Process the URL". https://worker.mturk.com/projects/36M7HST2M6WBKLBGK4TAQS9CXWX3B2/tasks?ref=w_pl_prvw

(b)

Workers will be awarded $0.08 for each URL they process.

(c)

There's no qualification, eligibility requirement, or restriction.

(d)

The allotted time for this task is 100 minutes. But I think I can do 5 items in an hour. The implied hourly rate is $0.4/hour.

(e)

This jod expires in 7 days (on Nov. 17th).

(f)

If 1 million people participated in the task and each of them did one item, this project would cost the HIT creator $80,000.

4. Kaggle open calls

The competition "Quora Insincere Questions Classification"(https://www.kaggle.com/c/quora-insincere-questions-classification#description) is held by Quora, a knowledge sharing platform where people can ask questions, provide insights to other people's questions, and discuss with people who share the same interest. Quora is trying to detect "toxic and divisive" questions to make the platform a better place so they lauch this competetion to find "more scalable methods".

All submissions should use Kaggle Kernels. The final submission file should be a file named "submission.csv" containing two columns: "qid in the test set" and the corresponding 0/1 prediction indicating whether the question is "insincere". The file should contain headers "qid" and "prediction", and the fields should be seperated by a comma. This file should be commited to the Kernel and then the Kernel should be submitted to competition by clicking the "Submit to Competition" button in the output tab. Participants should be aware that their submissions should not have internet access, multiple data sources, or custom packages. Submissions are evaluated by F1 score, which is the harmonic mean of the precision and recall rate between the

observations and predicted values.

The timeline of this competition is as follows: Jan. 29, 2019 is the "entry deadline", by which participants should "accept the competition rules"; Jan. 29, 2019 is also the "Team Merger deadline", by which the team merging process should be done; Feb. 5, 2019 is the "Final submission deadline".

Finally, the 1st place winner of this competition will be awarded $12,000, and $8,000 for the 2nd place winner, $5,000 for the 3rd place winner.