hw6 Problem2&3&4

November 13, 2018

1 Assignment 6

1.0.1 MACS 30000, Dr. Evans

1.0.2 Zeyu Xu

Due Monday, Nov. 19 at 11:30 AM

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

See the attached pdf.

1.0.4 2. Collaborative problem solving: Project Euler

(a) Here is my Project Euler user name and friend key:

zeyuxu1997: 1408161_9ckXcwM2BApMN4pu5UHuXk80sGZDcHAt

(b) Here is the code to solve Problem 1: Multiples of 3 and 5.

```
In [1]: Sum = 0
        for i in range(1000):
            if i % 3 == 0 or i % 5 == 0:
                Sum += i
        print(Sum)

233168
```

As shown above, the answer to this question is 233168

(c) I would most like to achieve D for Dedication, As Easy As Pi and Big Game Hunter. D for Dedication and Big Game Hunter are all challenging, the former one is a challenge both to patience and intelligence and the latter is extremely challenging for it required me to solve most difficult problems. As Easy As Pi is interesting because it requires me to complete problem 3, 14, 15, 92, 65, etc, which is coincident with (3.14159265...).

1.0.5 3. Human computation projects on Amazon Mechanical Turk

(a) The name of the task is Translate Swahili to English. It asks workers to translate the body of the document found at http://mtanzania.co.tz/jpm-atema-nyongo/ from Swahili to American English. Word count is approximately 330 words in the document body including title and author affiliation.

(b) The payment is 5 dollars for each HIT.

(c) It has 1 requirements: Location is TZ.

(d) This job supposed to take 300 minutes (5 hours). So the implied hours rate is 1 dollar per hour.

(e) It will expire at 07:04 am on 11/16/2018.

(f) For one participant, the payment is 5 dollars, so if 1 million people participated in the task, it could take the HIT experiment creator 5,000,000 dollars at most.

1.0.6 4. Kaggle Open Calls

(b) The title of the competition is Two Sigma: Using News to Predict Stock Movements. The sponsor is Thomson Reuters. It's a multinational mass media and information firm. The evaluation rule is like that: participants predict the returns of a stock for 10 days and assign a confidenceValue to the stock (positive for positive returns and negative for negative returns, the absolute value represents both the participant's confidence on the prediction and the scale of returns), and the sponsor will calculate the participant's score with a formula using the data of the stock in the evaluation period, the more accurate the prediction, the higher the score, and finally take the average score as the final score. The prize structure is 25,000 dollars for the 1st place, 20,000 dollars for the 2nd place, 15,000 dollars for the 3rd place and 10,000 dollars for the 4th to 7th place. The honor code is that participants are forbidden from using information that is outside of the provided dataset, or using the provided information in a way that is not intended. The sponsor maintains the right to disqualify any contributions that they believe demonstrating cheating. The competition is a two-stage competition consisting of a Submission period and a Scoring period. It starts on 9/25/2018 and end on 7/15/2019. The deadline for rules acceptance and team merger is 1/2/2019. The deadline for Submission is 1/8/2019. Participants can have many submissions and choose 2 best ones to be rescored before the deadline. The Scoring period is a period when models submitted are evaluated against regularly updated news and market data. The submission instructions can be summarized as follows: All submissions will occur through the Kernels environment. Participants must use python module to access the competition data, step through time, collect predictions, and write an appropriate submission file. The submission file should consist of both the historical, stage 1 time period and the future, stage 2 time period and thus there is only one "valid" submission file at a given time. Predicted values for the times outside of the current stage will not affect scores. The limits of constraints are (i)16 GB Memory; (ii)6 Hours total runtime; (iii)4 CPU cores.

(c) In my view, since the sponsor is a mass media and information firm, they will take the winning submission as reference to construct or improve their model to predict returns of stocks so that they can make better investment decisions or provide prediction service to customers.