

Problem Set #6

MACS 30000, Dr. Evans

Nan Ge

Question 1. Netflix Prize and Bell [Bell et al. \(2010\)](#)

(a) Describe how submissions to the Netflix Prize open call contest would be judged? That is, what was the criterion function? Were there any cutoffs beyond which a submission would not be judged (i.e., the t was so poor that it would be called a zero)?

Netflix handed out a training set of over 100 million ratings from almost 500, 000 customers. The participating teams were expected to develop a collaborative filtering model to predict individual ratings for a hold-out set. ([Bell et al., 2010](#), pp. 2) The predictions were evaluated by calculating their root mean squared errors (RMSEs) with regard to the real customers' ratings. A submission had to improve RMSE by 10% over Netflix's own algorithm, Cinematch, to be considered valid. The jury would only start grading submissions 30 days after the first team that has a 10% improvement appeared. ([Bell et al., 2010](#), pp. 7)

(b) At the beginning of the Netflix Prize contest, what was the most commonly used method for predicting ratings (stars) on movies?

At the beginning of the contest, the most commonly used method was nearest neighbors. It uses weighted average of ratings of similar items from a customer's rating history to generate rating predictions for a new item. ([Bell et al., 2010](#), pp. 3)

(c) The best predictive models in the Netflix Prize open call were hybrids of multiple models (ensemble methods). What characteristic of one model relative to other models made it improve the overall prediction when blended with the other models?

A model would improve the combined result if its input is uncorrelated or not highly correlated with other models. ([Bell et al., 2010](#), pp. 6) That is, adding in a new model would achieve an incremental reduction in overall RMSEs if it captures additional information of the input variables.

Question 2. Collaborative problem solving: Project Euler

(a) Register as a user of Project Euler. Report your Project Euler user name and friend key.

jsgean: 1407387_LYiueHgMTMewlsy2rX5WNvNObwQceOlm

(b) Look through the Project Euler archives of problems. The earlier problems are easier problems. Choose one of the problems and complete it using either Python or R programming languages. Report both your code and your answer.

Task: 1000-digit Fibonacci Number¹

```
def fib():
    first = 1
    second = 1
    index = 2
    while len(str(second))<1000:
        third = first+second
        first = second
        second = third
        index+=1
    return(index)
print(fib())
```

Output: 4782

(c) Look through the Project Euler Progress page. List the three awards that you would most aspire to achieving and describe what you like about those awards.

"Fibonacci Fever": Solve the first twelve Fibonacci numbered problems. I like this award because it combines accomplishment with fun.

"One In A Hundred": Be among the first hundred to solve a problem. I like this award because it speaks for my initiative and creativeness to solve problems.

"Wise Words": Received five kudos across your permanent posts. I like this award because it suggests I'm not only good at solving problems, but also a valuable contributor to the learning community. ²

Question 3. Human computation projects on Amazon Mechanical Turk

(a) Select an MTurk human intelligence task (HITs) that is a human computation project and IS NOT a survey or an experiment. Most HITs on MTurk are human computation project

The project is called "Find info from an email v1".

(b) Describe the full payment structure of this HIT. That is, the reward column says an amount, but there is a lot more information available as to what that amount means.

The task pays \$0.03 for every verified entry. The emails are being sent to multiple workers for doublecheck. Each worker is allowed 20 wrong/missing out of 100 emails.

(c) Describe any qualifications, eligibility requirements, or restrictions (or lack thereof).

There are two qualifications. First, HIT approval rate no less than 85. Second, total

¹<https://projecteuler.net/problem=25>

²<https://projecteuler.net/progress>

approved HIT is greater than 1000.

(d) What is the allotted time for this task? How many items do you think you could do in an hour? What is the implied hourly rate (dollars per hour)?

The allotted time is 60 minutes. I think I can process 200 emails in an hour. That makes the implied hourly rate about \$6 per hour.

(e) When does this job expire?

The job expires on Nov 16, 2019.

(f) What is the most this project would cost the HIT creator if 1 million people participated in the task?

I think there is no limit as to how many emails each worker can process, so maybe the requester just wants a dataset as big as possible. Suppose every worker works for exactly one hour, at my speculated entry rate, and their answers are all approved, the maximum cost for this project would be 6 million dollars.

Question 4. Kaggle Open Calls.

(a) Register for a Kaggle account from the Kaggle home page.

username: jsgenan

(b) Describe one of the open competition. Make sure that your description is paraphrased (in your own words) and not just copied and pasted from the text in the open call project. Include in this description the following information:

- The title of the competition
- The sponsor of the competition
- A description of what type of company or what type of person the
- Sponsor of the project is
- How submissions will be evaluated
- Prize structure for winning submissions
- Any honor code issues of importance
- Timeline description
- Submission instructions

One of the competitions that I'm particularly interested in is "Quora Insincere Question Classification".³ It is sponsored by Quora, the largest online community for

³<https://www.kaggle.com/c/quora-insincere-questions-classification>

questions and answers. According to its website, Quora aspires to "share and grow the world's knowledge".⁴ I like and use Quora a lot, for there are plenty of interesting questions, and the community always offers fresh perspectives to the same question. One key component to the atmosphere of the community is the quality of questions. Despite plenty of good questions, there are also many meaningless, insincere, even toxic questions on the forum. That's why Quora initiated this challenge to ask Kaggle teams to help them filter out these insincere questions using machine learning models. The participants are expected to use the training set and give a 0/1 prediction to whether a question is insincere. The submissions will be run by the jury on a new test set, and will be evaluated based on F-1 score between the predicted outcomes and the observed targets. Quora offers three different levels of prize: \$12000 for the first place, \$8000 for the second place, and \$5000 for the third place. There is no specific honor code, but the general competition rule forbids private sharing of codes. Also, no external data is allowed in this competition, except for the word embeddings provided by Quora. The entry deadline is January 29, 2019. A participant must accept the competition rules before this date in order to compete. Team merger deadline is January 29, 2019. This is the last day participants may join or merge teams. And the final submission deadline is February 5, 2019. This is a kernel-only competition. Each team could only submit five entries in one day. The results have to be written and saved in a .csv file named submission.csv. The script must disable internet access. Multiple data sources and custom packages are forbidden. Plus, the running time must be less than 6 hours for a CPU kernel, and less than 2 hours for a GPU kernel. Winners of the competition are obliged to deliver the final code to the sponsor.

(c) Given your answer about what the sponsoring entity does and your description of this project, what do you think the sponsoring entity will do with the winning submission answer? How will they use it?

I think Quora will incorporate the algorithm into their platform to filter out insincere, toxic questions, in order to create a better atmosphere in their community. While it's not probable to replace manual review team at the moment, machine learning is a much more efficient resolution to this problem.

References

Bell, Robert M, Yehuda Koren, and Chris Volinsky, "All together now: A perspective on the netflix prize," *Chance*, 2010, 23 (1), 24–29.

⁴<https://www.quora.com/about>