Perspectives in Comptational Analysis

Problem Set 6

Problem 1

(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 fit was so poor that it would be called a zero)?

The submissions would be judged on the ability of the submitted algorithm to improve accuracy in its website recommendation system over the current one (CineMatch) by at least 10%. More specifically, it had to reduce the Root Mean Square Error of the predicted rating on the test data set to 90% of the last achieved one (0.9525), which amounts to 0.8572. This, in effect, became the cut-off for consideration for the ultimate grand prize. (Netflix, 2007)

Till this 10% figure was reached, Netflix handed out a Progress Prize to the team that had made the maximum progress since the last year's highest improvement. However, the submission had to achieve at least a 1% over the last highest improvement to be considered a valid entry.

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

The most commonly used method for predicting ratings of movies before the Netflix Prize Contest was the collaborative filtering method called Nearest Neighbours. Essentially, a user's predicted star rating for an item would be calculated as the weighted average of his/her ratings for similar items.

(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?

The authors state that as such, adding a new model to an existing model or ensemble thereof did not reduce the overall prediction accuracy. There would exist some linear combination of the k+1 models that raise the accuracy over that original ensemble of k models.

In this specific case, the Netflix models were heterogeneous in that they were not highly correlated and 'did not share anything'. According to Smolyakov (2017), the individual members of a heterogeneous ensemble of methods individually need to be internally

accurate and methodologically distinct from each other for the ensemble to outperform its constituents.

Problem 2

(a) Register as a user of Project Euler. I put my Project Euler friend key (below) on the last page of my CV. Report your Project Euler user name and friend key.

policyglot: 1408690_phBx91MvzoEWusvGZKXsEbSGzA2cM0uK

(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.

I selected Problem 1.
https://projecteuler.net/problem=1

The code is available as a Jupyter Notebook in the same folder as this file.

The answer was 233168.

(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.

I would most aspire to earn the 'The Guru', 'Decathlete' and 'Triangle Trophy' awards.

The 'Decathlete' appeals to me because it resonates with personal interests. I aim to become a triathlete in the next 5 years. Project Euler doesn't have a Triathlete award, but 'Decathlete' comes close.

The 'Guru' is a word of Indian origin, much like me. In addition, it is granted for contributions to writing pieces for the website. I consider myself to be a much better thinker and writer than a coder per se. So an award of this kind would highlight my strengths

The 'Traingle Trophy' caught my eye primarily because of the visual appeal of the logo. Furthermore, the requirement to win it psychologically strikes a chord as it's challenging without being overwhelming.

These awards thus demonstrate the use of a number of facets used in gamification- such as appealing to the users' sense of identity, self-perception and psychological needs for challenge and superiority over others.

Problem 3

(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 projects.

I selected the task 'Transcribe up to 35 Seconds of Media to Text - Earn up to $0.17 per HIT!!' posted by Crowdsurf Support.

(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 base rate per HIT submission is $0.05, with up to $0.17 possible with the addition of reward amounts based on the length of the transcription (where the maximum is 35 seconds).

Within the sample HIT provided, the task mentions 'Happy Hours' on Sunday, 11/18/18 from 8:15am - 11:59pm PST. During this time, each approved FTR and FTR-GM task receives an additional $0.06 per mm (media minute). This policy applies to those workers who have a minimum QC (Quality Control) score of 3.80.

In addition, a weekly transcription bonus is available for those who have who have a minimum QC score of 3.85 and our considered 'Level 3 and Up'. The worker must transcribe at least 100 mm and maintain a QC of 3.8 or more. Since the maximum time is about 30 seconds, such a bonus would require about 200 tasks in a week.

Between the 'prorate' and bonus rates, only the highest rate (not the cumulative amount) will be paid. Bonus payments become available starting on November 18, 2018. In addition, no bonus or regular rates will be paid if fraud is detected.

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

There are two qualifications/restrictions for this task.

First, the worker needs to have a Human Intelligence Task approval rate of 95%. Since I have a new account with no tasks yet, my score by default is 100%. Hence, the first requirement is looked after.

The second eligibility requirement is 'Qualified to work on Transcription Tasks is not less than 900'. Here, the worker needs to take a one-time test, whose link is provided just below the task description.

(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 15 minutes per HIT submission. On the basis of the sample I looked at, it would take me at most 2 minutes to transcribe the recorded media segment. So I could complete about 30 such tasks in an hour. This would imply an hourly rate of 30*0.05= $1.50. If I received the highest possible reward on each task, the pay would rise to $0.17 per task. In such a case, my hourly rate could rise to 30*0.17, or $5.10.

(e) When does this job expire?

The task expires on the 17th of November 2019 (roughly a year from the present date

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

If each of the million participants achieved the maximum reward of $0.17, the creator would need to pay a total of $170,000.

Problem 4

(a) Register for a Kaggle account from the Kaggle home page.
I have registered with the username 'policyglot'

(b) Describe one of the open competitions. 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

This competition, entitled 'Costa Rican Household Poverty Level Prediction', is sponsored by the Inter-American Development Bank (IDB), which describes itself here as 'the largest source of development financing for Latin America and the Caribbean'. It is a multilateral international institution owned by 48 sovereign states (IADB, 2018). The purpose of this competition is to develop algorithms to correctly classify households as per their poverty levels and accurately predicting their level of social need. Participants would be expected to extend their approach beyond conventional econometric methods, including the IDB's currently used PMT ('Proxy Means Test').
https://www.kaggle.com/c/costa-rican-household-poverty-prediction

Submissions would be judged based on the final 'Macro F1 Score', which is the harmonic mean of the test's precision and recall (Yutaka, 2007). In this specific case, since there are 4 categories in the classification problem, we will be taking the average of the four individual F1 scores for each category (Koehrsen, 2018). This averaging process is what led to the addition of the term 'Macro' before 'F1 Score'. However, exploratory data analysis (based on only the heads of households, and not all the individuals in the data) reveals that the category 'Non-Vulnerable' is over-represented. By this logic, a weighted F1 score may have been more suitable for this task. (ibid).

Participants would be expected to submit their 'kernels', which, as per Kaggle's new definition, refers to "combination of environment, input, code, and output - all stored together" (Montoya, 2016). The code within the kernel can be shared as well as edited by team members on the Kaggle platform.

Prizes are being awarded broadly in categories- 'Top Kernel Prizes' and 'Top 3 Forked and Submitted Prizes'. While authors of multiple kernels may win several awards, a given kernel itself can only be provided one award.

For the former category, submissions were assessed in two different periods of two months each, with 5 prizes awarded at the end of each period- specifically on the 17th of August and the 19th of September 2018 respectively. The judging criteria relates to the upvotes (excluding by the same user or by admins), and thus relates to the perceived quality as per other participants. The organizers seem to be aiming for inclusivity by declaring that awards in each period are bestowed on those who have never previously received any.

For the latter category, 3 prizes are awarded only once at the end of the overall competition time frame (see below). The judging criteria here is the number of forks of a participant's kernel and the unique number of submissions based on that fork. Here, the criterion appears to be not just perceived quality, but how a contribution helps build the foundation for future work by others. It could thus be argued that prize rewards collaboration and mutual learning even more explicitly than in the first category.

The contest began on the 18th of July, 2018. Its final deadline for submissions was set at 11:59 PM UTC on the 19th of September, 2018. The organizers have reserved the right to update the timeline as they see fit. The competition is still accepting entries using the 'late submissions' button.

Instead of the usual .csv files containing predictions, the participants are required to submit their kernels on the Kaggle platform. Within this Kernel, each ID must predict the appropriate class for the Target variable. This ID is only provided at the level of each household, though data may be available at the level of its individuals. These two are thus the only columns in the file. The description includes two sample rows in the file, along with the headers:

Id,Target
ID2f6873615,1
ID1c78846d2,2

While there is no clearly defined 'honour code' the rules of the competition prohibit pre-trained models, external data, or cheating (such as hand-labelling). The website clearly states that points from this competition do not count towards overall Kaggle ranking or medals.

(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?**
The Inter-American Development Bank describes among its focus areas "three development challenges – social inclusion and inequality, productivity and innovation, and economic integration" (IADB, 2018)

As suggested by their mandate, I believe they will use the kernel of the winning team in their own machine learning models and operations to correctly identify which members of the surveyed population require aid the most. While the precise use is not delineated on the competition page, the winning contributions could be applied for more effective implementation of its initiatives such as the Prospera Conditional Cash Transfers programme in Mexico, etc. (Davil-Larraga, 2016)

BIBLIOGRAPHY

Davil-Larraga (2016). 'How Does Prospera Work?' Inter-American Development Bank.
https://publications.iadb.org/bitstream/handle/11319/7569/How-does-prospera-work.PDF?sequence=4&isAllowed=y
(Accessed on Nov 18, 2018)

IADB (2018). 'How We Are Organized'. Inter-American Development Bank
https://www.iadb.org/en/about-us/how-are-we-organized
(Accessed on Nov 18, 2018)

Koehrsen, Will (2018). 'A Complete Introduction and Walkthrough. Kaggle.
https://www.kaggle.com/willkoehrsen/a-complete-introduction-and-walkthrough
(Accessed on Nov 18, 2018)

Montoya, Anna. (2016). Kaggle Kernel: A New Name for 'Scripts'.http://blog.kaggle.com/2016/07/08/kaggle-kernel-a-new-name-for-scripts/
(Accessed on Nov 18, 2018)

Netflix (2007). The NetFlix Prize Rules
https://www.netflixprize.com/rules.html
(Accessed on Nov 18, 2018)

Sasaki, Yutaka. (2007). The truth of the F-measure. Teach Tutor Mater.
https://www.researchgate.net/publication/268185911ThetruthoftheF-measure
(Accessed on Nov 18, 2018)

Smolyakov, Vadim. (2017) 'Ensemble Methods to Improve Machine Learning Results'. Medium.
https://blog.statsbot.co/ensemble-learning-d1dcd548e936
(Accessed on Nov 18, 2018)