Assignment #6

Sanja Miklin

11/19/2018

1. Netflix Prize and Bell, Koren, and Volinsky (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 fit was so poor that it would be called a zero)?

The contest submissions were judged by their ability to predict ratings on a test data-set, as compared to the Netflix's algorithm *Cinematch*. The measure of fit was the root-mean-square error (RMSE). In order to qualify for the prize, a submission had to improve predictions by 10% or more, that is the RMSE of the model had to be at least 10% lower than the Cinematch RMSE.

The winner of the contest was to be determined using a different held-out test-set. However, it turned out that the top two submissions had the same RMSE (rounded to 4 decimals), so the winner was decided by time of submission.

From the article, it is unclear whether other entries were judged at all (it would seem not). However, I was curious about the contest and browsed through the Netflix Prize Wikipedia Page, where I've discovered that Netflix also gave out a " 'dubious honors' (sic Netflix) of the worst RMSEs on the Quiz and Test data sets," so it would seem all the submission might have been judged.

(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 common method was 'nearest neighbors', which predicted ratings of a movie based on an average of the user's ratings of similar (neighboring) movies.

(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?

One model can improve prediction when blended with other models, if it is not highly correlated with other components. Apparently, such blending works to improve prediction even if the contributing model is quite inferior.

2. Collaborative problem solving: Project Euler

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

smiklin: 1407496_Kna8ru8TPZqPNmxBshDWnRRnAFjLH4jU

(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.

For a first attempt, I chose the problem *Even Fibonacci numbers*: Each new term in the Fibonacci sequence is generated by adding the previous two terms. By starting with 1 and 2, the first 10 terms will be: 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, ... By considering the terms in the Fibonacci sequence whose values do not exceed four million, find the sum of the even-valued terms.

```r
#set empty vector

fibonacci <- c(1,2)
i <- 2

#loop to get the fibonacci sequence
while (fibonacci[i] + fibonacci[i-1]  < 4000000){
    i <- i+1
    fibonacci[i] <- fibonacci[i-1] + fibonacci[i-2]
}

# sum even elements
sum(fibonacci[which(fibonacci %% 2 == 0)])
```

```
## [1] 4613732
```

The sum of the even elements of a Fibonacci sequence whose elements do not exceed 4 million is 4,613,732.

(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 really like the names of some of the awards, specifically 'High Flyer' (I am a flyer in circus), 'Master of Archives' (I love archival research) and 'As easy as Pi' (because I love pies and pi puns). Of those, I think the last one might be the most achievable, but we'll see (so far, I'm working towards the 'Decathlete,' it seems).

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 projects.

I selected a HIT titled *Rate the contents of a Tweet: 10 HITS MAXIMUM* posted by a Robert Meyer. The task is to rate 10 tweets along a number of dimensions.

(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 most I can see is the reward, which is written as $0.05

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

The qualifications are as follows 1) HIT approval rate (%) is greater than 97 2) Number of total approved HITs is greater than 50 and 3) Worker's location is US

(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 10 minutes, and (if allowed by the HIT) one could probably rate 60 (if not more) tweets an hour. The implies hourly rate is, therefore, $0.05 \times 6 = \$0.30$ per hour.

(e) When does this job expire?

The job expires on November 14th.

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

If 1 million people participated in the task, the project would cost the HIT creator $1.2(\$0.05 \times 1,000,000) = \$60,000$

4. Kaggle open calls

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

Done!

(b) Describe one of the open competition.

The competition that caught my eye is *Quora Insincere Questions Classification Detect toxic content to improve online conversations* because I am really interested in language/language processing as well as online discourse. The sponsor of the competition is Quora, a large private question-and-answer website that encourages participation under one's real name (it boasts contribution by well known figures such as \href{https://www.quora.com/profile/Barack-Obama-44}{Barack Obama or Stephen Fry) and is concerned about preserving the productive and honest nature of the questions.

The competition is looking for models that can identify and flag "insincere questions", such as those that are not actually looking for answers, but are intended to make some sort of a statement, or are posited under false premise. The first place prize is $12,000, second place prize is $8,000 and third place prize is $5,000.

Submissions to the competition have to be submitted through the Kaggle's Kernel system, as 'submission.csv', a set of question ids from the test set with the corresponding prediction ('0' if sincere, '1' if insincere). The submissions are evaluated based on how well the predictions match the observed (determined by Quora) values, through an F1 Score.

Participants can enter the competition and join/merge teams until January 29th, 2019. Submission deadline is February 5th, 2019.

(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?

It seems likely that Quora would integrate the winning model into the process it has been using (machine learning and manual review) to flag insincere questions.