THE DATA INCUBATOR ASSIGNMENT

Alvin Lawson

6/8/2018

Thank you for providing your availability. We were really impressed with your background! Sorry for the late response: we've been heads down on a few client projects recently. If you're still interested, we'd love to start the interview process. We're different than most companies: we emphasize project work to help you prepare for the interview, rather than just giving you brain teasers out of the blue. We want to see how well you're able to explain topics in statistics and data science. Write a short Jupyter notebook (and put it on github) covering these topics:

Question:

1. Hypothesis Testing: Let's talk about t-tests, p-values. How are they related? What is it telling you? How does it relate to precision-recall? What are the underlying assumptions?

Answer:

Hypothesis testing tests the outcomes of an experiment to see if there are any important results. When it comes to Hypothesis testing there are two terms that individuals need to know. The two terms individuals need to know is T-test and P-Values. A T- Test compares two averages to let you know if they are different from one another. On the other hand, P-Values checks to see the probability of a null hypothesis in the study question is true. T-tests and P-values are related since they are both trying to find out if the null hypothesis is true or not. This relates to precision-recall because it helps in measuring the success of a hypothesis when there are imbalanced classes. The underlying assumption is that precision helps in calculating relevancy. On the other hand, recall helps in calculating how many applicable results returned.

Question:

2. Bayesian posterior inference: Explain Bayes' Rule. Write some code to actually perform posterior sampling. Work out an example using conjugate priors. How does this compare with hypothesis testing? What are the underlying assumptions?

Answer:

Bayes Rule finds the probability of an event based on the previous information related from that event. This posterior sampling compares to hypothesis testing since we are making an educated guess to find the results of 20 tosses when observing 5 heads. Eventually helping us to compare it to the null Hypothesis. The underlying assumption is that we make 20 tosses and we observe 5 heads.

Be prepared to give a short mock “lecture” (30 minutes) about these two topics with your prepared Jupyter notebooks. The Jupyter notebooks are more meant to be notes for yourself and visual aids for your lecture.
We’ll be looking for

1. How well you present: remember that this material should be approachable, applied, and not just a series of formulas
2. How well you understand these topics in depth (the mathematics, the underlying assumptions behind ideas, etc ...)

In Addition:

1. Please provide sometimes that would be good for an interview later next week or the week after (if you need more time).

2. Send your GitHub link before the interview.
https://github.com/al11588/DataIncubatorAssignment

3.