

Math 514

Project Proposal

April 9, 2019

Project Proposal

Project proposals are due Wednesday, April 10. Project teams will give a short presentation during class on April 24. A set of slide templates using the Xaringan environment will be provided to help with preparation. Please make an effort to share what you have learned about your topic with the class. The final project is due by midnight May 5. The final deliverable should use reproducible research methods and must run successfully.

1. Identify team members. Include name and email

Our team consists of Mary Peng (xp19@georgetown.edu) and Michael Leibert(mll82@georgetown.edu)

2. Provide a problem statement and a description of what you hope to achieve. If you are starting with a p

We will use fine foods reviews to predict user sentiment toward products reviewed on Amazon. Our dataset contains both short review summary, as well as the full length review. We will start by focusing on predicting user sentiment from the short review summaries.

We are particularly interested in different regularization techniques, and will compare the effectiveness of dropout vs. L1 / L2 regularization.

3. Give a description of the data set The dataset comes from Stanford's SNAP database (<https://snap.stanford.edu/data/web-FineFoods.html>). The dataset contains 568,454 reviews by 256,059 between 1999 and 2012. We will split the dataset into half training and half test. Sample data format is as follows:

product/productId: asin, e.g. amazon.com/dp/B001E4KFG0 review/userId: id of the user, e.g. A3SGXH7AUHU8GW
review/profileName: name of the user review/helpfulness: fraction of users who found the review helpful
review/score: rating of the product review/time: time of the review (unix time) review/summary: review
summary review/text: text of the review