Daniel Conde

Michael Burks (Instructor)

UCB Bootcamp, MW 6:30 - 9:30 p.m., Sat 10 a.m. - 2 p.m.

16 October 2019

Final Project Proposal

Idea: Using Machine Learning to Predict Amazon Product Rating

Source(s)

<https://www.kaggle.com/bittlingmayer/amazonreviews>

<http://jmcauley.ucsd.edu/data/amazon/>

Summary

The goal of this project is to use basic knowledge of machine learning techniques and neural networks that were studied in class to develop a program that predicts the star rating of an Amazon product by utilizing sentiment analysis on customer reviews (and, if time allows, by comparing the product’s price to the average price of similar products and by taking into consideration the number of purchases for that particular product). Often, an Amazon product can become subject to a low rating for several reasons other than the product’s performance or quality, such as a poor shipping and handling experience on the consumer’s end, wrong sizing for clothing and apparel, or an inaccurate product description. So, I want to see if certain star ratings on Amazon products corresponds to specific outcomes experienced by the customer.

The data gathered will be primarily from the two sources above. The kaggle dataset contains comprehensive data on a variety of customer reviews, while the second source from Julian McAuley contains data on specific Amazon categories. Because both data sources have extremely large datasets, I will have to take pieces of data from each source manually as well as narrow the number of product categories that will be considered in this project.

Process

1. Gathering Data
2. Data Pre-processing / Cleaning the Data
3. Determining Model → Categorical and/or Regression Model
4. Training and Testing the Model
5. Project Evaluation

Projected Timeline

1. Saturday, October 19 → Gathering and parsing pieces of datasets
2. Monday, October 21 → Finish cleaning data and Implementing data into program
3. Wednesday, October 23 → Training data and testing data
4. Saturday, October 26 → Fixing bugs and then more training and testing
5. Monday, October 28 → Finalize model and evaluate results (and make a powerpoint)

Sorry this was late.