Predicting Advertising Revenue through Regression

Zoheb Siddiqui

October 21, 2018

0.1 Introduction

I work, part time, as a data science intern at an advertising company. My project, which will be a part of my work responsibilities, is to take in advertisement data for any random client that my company takes on and use it to predict revenue for that client given a budget and a fixed time frame to utilize that budget in.

0.2 Project Idea

The idea of the project is to create a python program that takes in the name of a client, the budget assigned to them and the time frame to advertise over. The output will be a table of advertisement features, user demographics and ad placements along with the optimal amount of budget each data point should receive and the revenue it will generate.

I have advertising data for 3 clients. This data is collected over a period of 3 months. It contains information regarding ad features, targeted demographic and ad placement.

One of the major concerns I need to take into account is the fact that each country/state has a limited number of people and so does each gender and age-group. This would be a limiting factor when advertising.

0.3 Software I Plan on Using

Although this list is not concrete as I haven't thought this project through yet, I shall be using Python, MySQL, numpy, pandas, matplotlib, scikit-learn, and possibly Keras on a Tensorflow back-end to implement my project.

0.4 Papers/Textbooks to Read

1) Hastie, Trevor, et al. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, 2017.
2) Perlich, Claudia Dalessandro, Brian Stitelman, O Raeder, T & Provost, F.(2013). Machine Learning for Targeted Display Advertising: Transfer Learning In Action. Machine Learning. 95. 10.1007/s10994-013-5375-2

0.5 Will I have a Project Partner?

I have 1 project partner: LANGLEY PAUL DEWITT

0.6 Milestone

By the milestone, I hope to have cleaned the data, come up with an algorithm to incorporate the limiting factors mentioned above and tested a few regression algorithms on the data I have while answering why some work better than others.