Computer Science 524: Introduction to Optimization Project Spec Fall 2022

Author: Yanni Zhang

Regression Problem on Facebook Ad Campaign—Project Spec

(a) What is the issue being addressed?

Companies spend money to put advertisements on Facebook to attract people make a purchase or do a download. So naturally, it would be interesting to analyze the relation between the actual conversion and potential predictor variables, so that they could later make more efficient marketing strategies.

(b) Where does the data come from and how will it be obtained?

I found the dataset from Kaggle[1]. It includes over 1000 ads displayed on Facebook in 2017. It covers in total 3 distinct ad campaign and reports the target age and gender group, number of clicks, total and approved conversion number and so on. They are all from anonymous organization's social media ad campaigns. As it contains a column named fb_campaign_id that associated with how Facebook tracks each campaign, I would consider they are authentic and representative data.

(c) What is the optimization problem underlying this project?

This is a clearly regression problem. I would use least squared method to minimize the residuals. One interesting point would be considering quadratic or any other nonlinear form of the function. I think another possibility is that this could also be an optimal tradeoffs problem, as I want maximize the total and approved conversion rate but also minimize the total spending on putting all the ads. So doing cross-validation using different tradeoff parameter would be another part.

(d) What are the deliverables?

I would first find appropriate predictor variables, as some of them are categorical variables rather than representing real number. Then, I would formulate the model to find βs of predictor variables by adjusting the form of the function using least squares estimation. Next, I would use cross-validation to find a proper tradeoff parameter, but as I don't have more test data, I would randomly select some data point to be the 'test set'. Finally, if possible, I would draw the Pareto curve using pandas in python.

(e) Other points to consider when evaluating.

For now, I would be concerned about that whether I could actually incorporate those categorical variables into my model, or just be able to use other numeric columns.

[1] https://www.kaggle.com/datasets/madislemsalu/facebook-ad-campaign?resource=download