Dentsu Aegis Data Event

Ryerson Applied Math M.Sc Group

January 30, 2020

Contents

Chapter 1

Preliminary Thoughts

1.1 Introduction

I set this up for us to have a place to write any notes before going into the event and during so we can share any helpful tool and resources we have. From the little reading I've done it seems like we'll be looking at an online marketing exercise. Most of the services they've provide have been using AI and ML algorithms in order to provide targeted ads (eg. using face recognition on social media to determine which people have pets and push pet food ads through their feed).

- Alin

1.2 Logistic Regression

A very well known quantity in online marketing campaigns is logistic regression. It uses the logistic function in order to determine the likelihood of success (ad being appropriate for a given person) and then based on a threshold makes a binary decision on whether or not you should push the ad out to the user. This can be extended to a multinomial logistic regression where we can compute likelihoods for multiple categories (use threshold to say yes, maybe, no) on the ad and then perform a constraint optimization factoring in costs of pushing ads.

The model is set up as follows.

- Let $p = \mathbb{P}(X = 1)$ be the probability of a success

- Let $\{x_1, \ldots, x_n\}$ be the covariates

- Let $\beta_0, \beta_1, \ldots, \beta_n$ be the coefficients

- note that logit $: \mathbb{R} \to (0,1)$ is defined as $\text{logit}(x) = \frac{1}{1-e^{-x}}$

With the regression equation being

$$\text{logit}(p) = \beta_o + \beta_1 x_1 + \ldots \beta_n x_n \tag{1.1}$$

Fitting of this model can be done in several ways with the most common being the maximum likelihood estimator (MLE). An example of logistic regression in R via the `glm` function along with several tests for assessing the quality of the fit.

1.3   Possible "Solutions"

Below are two standard applications of ML in digital marketing. Several algorithms can be used to solve these problems which will be detailed as needed but in the simplest case can be solved with the logistic regression model above.

1.3.1   Content Curation

This problem is one of filtering. Provided with some information about a user, we should be able to eliminate "annoying" content/ads from their experience while at the same time providing ads that are most likely to drive business.

1.3.2   Product Recommendation

In simple terms, this is a matching exercise. Provided with some information about a client, we should be able to offer some products that they are more likely to be interested in. Solving this problem can be done using a multinomial logistic regression where the dimension of p (the vector containing probabilities of "success") is equal to the number of items in your inventory. The model should be trained on previous purchasing data. It is worth noting that majority of the covariates will be factor variables so regression may appear to be a poor fit.

Chapter 2

Data Work

2.1 The Dataset

2.2 EDA

2.3 Insight and Solutions