CS229 Project Proposal
Uplift Modeling : Predicting incremental gains

Akshay Kumar
akshayk@stanford.edu

Rishabh Kumar
rkumar92@stanford.edu

October 2018

Uplift modelling is a predictive response modelling technique which models the "incremental" effect of a treatment on a target group. For example, it has been applied to political election and ad products to study the effective of a political campaign and ad effectiveness. Traditional response modelling techniques build a predictive model to predict the response of an individual to a treatment (for e.g., seeing an ad campaign) based on prior response of treated individuals. In contrast, uplift model predicts the response of a treatment based both on treated and control population.

Uplift modelling is unique in the sense that it is composed of two different factors: let's assume we are interested in the effect ad on the purchase of a particular product. Uplift modelling is interested in the modelling what "additional" purchases an ad brings in i.e. $P(\text{purchase}|\text{treatment}) - P(\text{purchase}|\text{no treatment})$. Here, treatment implies watching the ad and no treatment implies not watching the ad.

Major techniques used for uplift modelling can broadly be classified into two categories: tree based methods and regression analysis. Tree based method construct a random forest by doing binary splits based on metrics like gini impurity. Regression based methods mostly stem from logistic regression to determine the probability of uplift. The simplest way to do this is to construct two different models: one to determine the probability of purchase with treatment and one without treatment and simply take the difference of the twos. However, the analysis of fitting the difference of two models is mathematically complex if the two individual models are imperfect. Another approach proposed in [Lo02] is to split each input feature into three: one corresponding to main effect, one to treatment effect and one to uplift effect. The final result then comes just from one model.

In this project, we will apply uplift modelling on Hillstrom email dataset [hil]. This dataset contains a list of emails for an ad campaign for an online retailer. The result of the email campaign was tracked for two weeks after the email campaign was over. Overall, there are 64,000 customers with some purchases in the past twelve months. They divided this population into three groups: one-third received a mail featuring men's merchandize, other one-third featuring women's merchandize and the remaining one third received no email. Each record in the dataset contains 8 features and finally 3 outputs: one each corresponding to visit, conversion and spend.

We will predict it using uplift modelling by a logistic regression model. The logistic regression model will be trained both on the control data (with no email campaign) and treatment data (where an email campaign was sent). Finally, we will also try a layered approach for uplift prediction as described in [Lo02].

To evaluate the performance of our model, we will examine qini curve as shown in Figure 1. To compute qini coefficient, we first sort (in descending order) each of the individual based on their predicted uplift scores. We plot this against the actual uplift achieved. If the model performs optimally, it should follow red line i.e. assign highest score to the individuals having highest incremental value. A random model would follow the diagonal line. Qini coefficient is simply the area of the model's curve above diagonal to the area of the optimal curve (*i.e.* red curve) above diagonal.

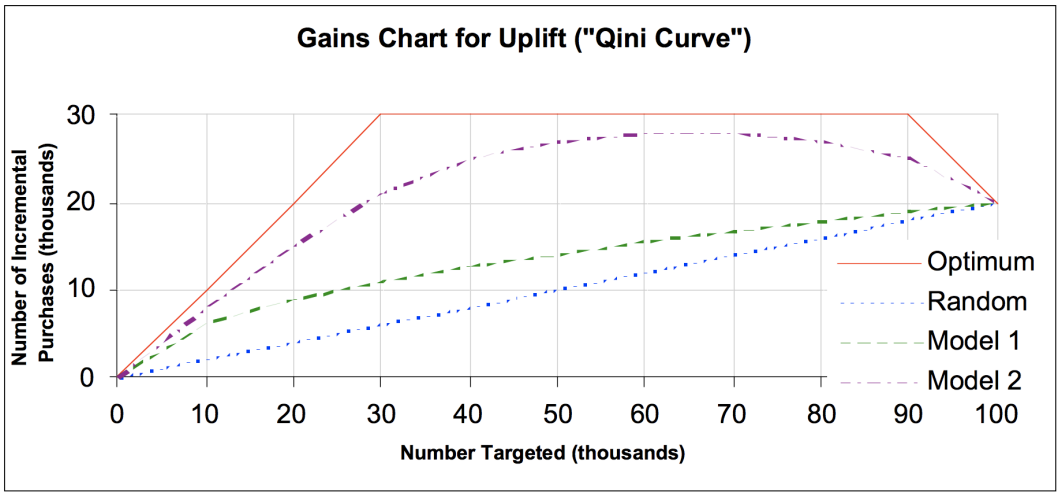

Figure 1: Qini curve [Rad07]

References

[hil] Kevin hillstrom: Minethatdata. https://blog.minethatdata.com/2008/03/minethatdata-e-mail-analytics-and-data.html. Accessed: 2008-03-20.

[Lo02] Victor SY Lo. The true lift model: a novel data mining approach to response modeling in database marketing. *ACM SIGKDD Explorations Newsletter*, 4(2):78–86, 2002.

[Rad07] NJ Radcliffe. Using control groups to target on predicted lift: Building and assessing uplift models. *Direct Market J Direct Market Assoc Anal Council*, 1:14–21, 2007.