Uplift Modeling : Predicting incremental gains

Akshay Kumar, Rishabh Kumar
CS229 | Stanford University | 12/11/2018

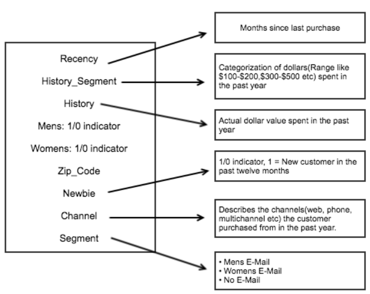
Introduction and Motivation

- **Uplift modelling:** predictive response modelling technique which models the "incremental" effect of a treatment on a target group.
- Traditional response modelling techniques just look at treatment group.
- $P[\text{purchase} \mid \text{treatment}] - P[\text{purchase} \mid \text{no treatment}]$
- In this project, we model the uplift modelling for certain email campaign for an online retailer i.e what "additiona" purchases an email campaign brings in for the product.

Dataset and Features

Hillstrom email dataset

- Email campaign related data for 64k customers with some purchase in past 12 months.
- Overall population divided into three different groups of equal size:
  - Received a mail featuring men's merchandize.
  - Received a mail featuring women's merchandize.
  - Received no advertizing mail.
- Each record contains total 9 features.
- Indicator variables indicating visit, conversion and spend.



Feature embedding & Algorithm Used

- Dataset has categorical features like segment, history_segment, channel etc.
- Inspired from word embedding in NLP, created one hot vector representation for each of these feature.
  - Learn different weights for each enum value.

Tackled the problem from two different perspective:

- Predictive response modelling
  - Also did ablative analysis
- Uplift modelling

Predictive Response Modelling

Experimented with the following configurations:

- Logistic Regression Model : FC followed by sigmoid activation.
- 3 Layer neural net : FC followed by ReLU followed by FC followed by ReLU followed by FC followed by sigmoid activation.
- Logistic Regression with bagging (Same as first but with bagging)
- Decision Trees : since many feature were based on enum values.

Training Config:

- Adam optimizer (gave better results than gradient descent optimizer)
- Loss function : cross entropy
- Mini batch gradient descent with batch size of 32.
- Trained the model for 5 epochs.

Also performed ablative analysis to get the most influential feature.

Uplift Modelling

Modelling "incremental" ad effectiveness.

- **Problem:** One individual training data : a user either sees an email campaign or do not see it.
- **Solution:** Two different models :
  - When no email campaign was seen.
  - When an email campaign was seen.
- Probability of purchase = Difference of the two models' predictions.

Uplift Modelling : Evaluation

Test data consists of points which either saw an email campaign or didn't see an email campaign.

- **Problem:** No definite labels for test data:
  - A single test data can not have both seen the email campaign and not seen the email campaign as well.
- **Solution:** Bucketization
  - Group test data with similar features into a single bucket.
  - Actual average uplift rate : Based on ground truth of labels for test data in the same bucket.
  - Compare actual average uplift rate v/s predicted uplift rate.

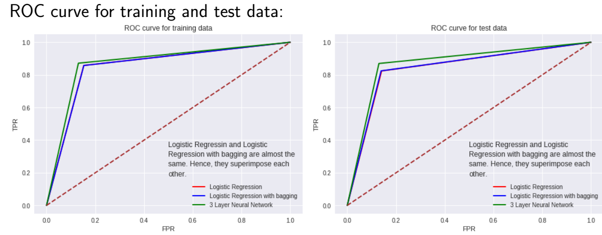
  Evaluation Metrics:
  - Qini Curve : Area under uplift curve.
  - Does not model negative uplift problem.

Results and Analysis

Predictive Response Modelling:
We split the whole data into 80% training and 20% test data. Since we have a class imbalance problem, we have to use a metric that is not biased towards the majority class. Therefore we have chosen to use F-score.

Model	F-Score(train)	F-Score(test)
LR	0.753	0.7313
BBLR	0.7689	0.749
3NN	0.801	0.79
Decision Tree	0.7129	0.6366

ROC curve for training and test data:



Ablative Analysis Results

- Evaluated importance of various features in logistic regression model by measuring drop in accuracy by dropping individual features.
- Did this on 3NN.

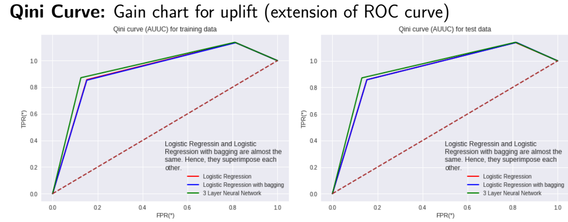
Results:

- Recency: 82.40 %
- History segment: 81.29 %
- History: 80.28 %
- Mens: 78.05 %
- Womens: 84.56 %
- Zip Code: 84.71 %
- Newbie: 84.62 %
- Channel: 85.14 %

Most powerful signal: Men's mechanize purchase in past 12 months.

Uplift Modelling : Results

Qini Curve: Gain chart for uplift (extension of ROC curve)



Conclusion

- We experimented with 4 different models for predictive response and neural network gave best f-score out of 4 models. Decision tree overfits the training data and predict poorly on test set.
- During uplift modelling, we can clearly see from Qini curve, uplift increases as we increase the treatment but decrease thereafter implying the possibility of negative effect on certain groups.
- Results of uplift modelling illustrates the possibility of achieving more incremental effect by targeting a smaller group.

Acknowledgement

We are thankful to Prof. Andrew Ng, Prof. Ron Dror and the entire course staff of CS229.