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.
- \[ P[purchase | treatment] - P[purchase | no treatment] \]
- In this project, we model the uplift modelling for certain email campaign for an online retailer i.e what “additional” 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 merchandise.
  - Received a mail featuring women’s merchandise.
  - Received no advertising 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.
- **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.

<table>
<thead>
<tr>
<th>Model</th>
<th>F-Score(train)</th>
<th>F-Score(test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>0.763</td>
<td>0.731</td>
</tr>
<tr>
<td>BBLR</td>
<td>0.7689</td>
<td>0.749</td>
</tr>
<tr>
<td>3NN</td>
<td>0.801</td>
<td>0.779</td>
</tr>
<tr>
<td>Decision Free</td>
<td>0.7120</td>
<td>0.6366</td>
</tr>
</tbody>
</table>

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.