Estimating the Causal Effect of Marketing Campaigns Using Doubly Robust and Tree-Based Methods

Susan Cherry
susan.cherry@duke.edu
STA 571: Advanced Machine Learning
Duke University

Motivation

This project investigates the effects of a direct marketing campaign conducted by a Portuguese bank. Specifically, it looks at the number of calls that the bank made to consumers and estimates the effect of the frequency of calls has on consumers' likelihood of subscribing to a term deposit. The question of how often to contact customers is of interest to many businesses. One one hand, frequent contact might increase the likelihood that a consumer buys a product but it could also have a treatment effect cap.

In general, casual effects are measured by comparing the treatment and control groups. A naive comparison of outcomes between these two groups is biased, since treatment is highly correlated with user features such as demographic information. Instead, causal inference attempts to model treatment as a random variable that depends on a set of pre-treatment covariates.

Traditional approaches to estimating the effect of a policy such as a marketing campaign include the direct method and inverse propensity scoring. Unfortunately, theory suggests that the direct method is prone to bias and that inverse propensity scoring often suffers from large variance. This project investigates the effectiveness of two newer approaches that are intended to reduce bias and variance: the doubly robust method and the tree-based method.

Data

The data used comes from the 'Bank Marketing Dataset" from the UCI Machine Learning Repository. It contains 41,188 observations gathered from a direct marketing campaign of a Portuguese banking institution. 30% of these observations are used for training. The remaining 70% are split into 7 test datasets. The methods are evaluated on each of the test datasets in order to determine variance.

The outcome variable is whether the client subscribed a term deposit. The treatment is the number of calls that the bank made to the client: 1, 2, 3, 4, 5, 6, or 7+. Features included in the model are age, marital status, education, employment, loan status, and whether the client had been contacted by a previous campaign.

Methodology

First, this project compares the naïve, inverse propensity weighting, and doubly robust methods. I train these methods on the training data and then evaluate on each of the test datasets. The mean and the variance of each treatment are recorded.

- **Naïve Method:** Estimate the treatment effect as the success rate of the various contact frequencies.
- **Inverse Propensity Weighting:** Can improve efficiency and reduce bias. The goal is to estimate the outcome that would be observed if each individual were assigned a certain treatment. Then compute the outcome if all individuals were assigned that treatment.
- **Doubly Robust Method:** Combines both a regression based estimator and the inverse probability weighted estimator. As long as either the propensity or outcome model is correctly specified, the doubly robust method is unbiased. Theory suggests this method will have a lower variance.

Next, I apply a tree-based method to the entire dataset. The tree-based framework is robust to misspecification and is highly flexible. The method is computationally efficient and unbiased.

The algorithm is as follows:

Input: Y_i, $\mathbf{X_i}$, treatment T_i for i = 1, 2, ...N.

Output: Estimated treatment effect.

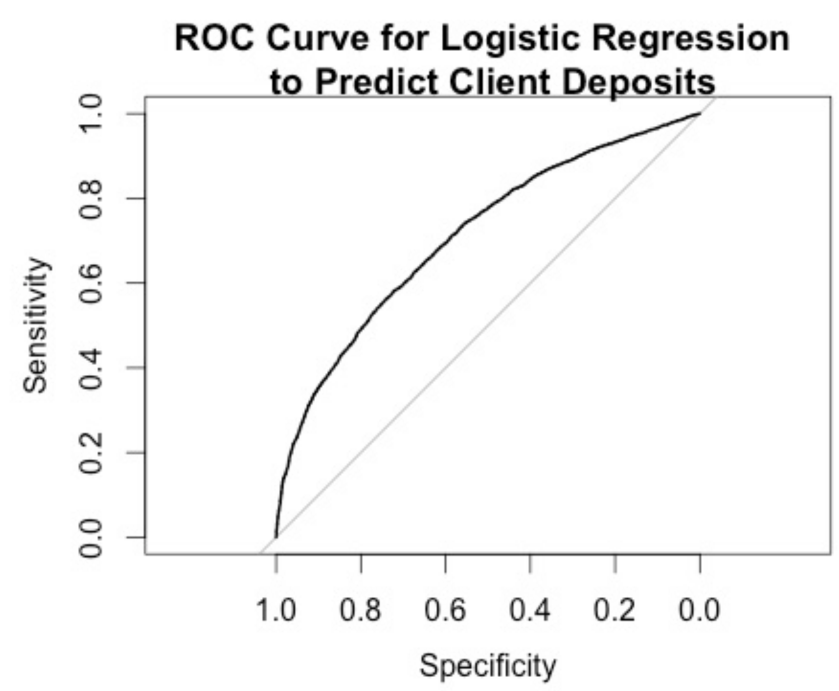
Step 1: Fit a tree-based model with dependent variable T_i and independent variables $\mathbf{X_i}$

Step 2: Within each leaf node, calculate the number of subjects and estimate the treatment effect for each treatment t.

Step 3: Calculate the final treatment effect.

Results

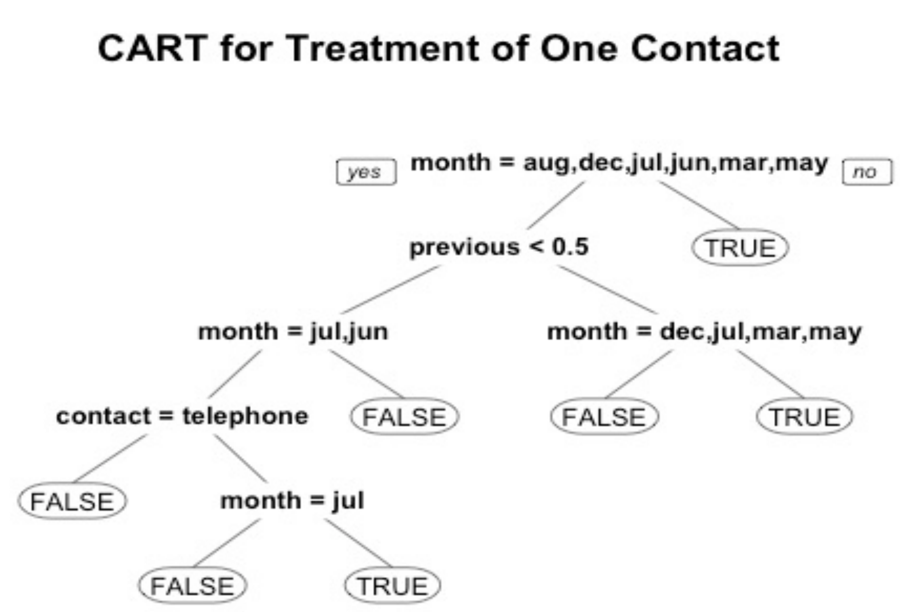
First, the ROC curve for the logistic regression used to calculated the probability that a client will subscribe a deposit is presented. These estimated probabilities are used in the inverse propensity weighting and doubly robust methods. The AUC is 0.734.



The table below contains the means and variances of the estimated effects from the naïve method, inverse propensity weighting method, and doubly robust method.

	Naive Method		Inverse Propensity		Doubly Robust	
# of Contacts	Mean	Variance	Mean	Variance	Mean	Variance
1	0.13046	0.00003	0.12326	0.00003	0.10039	0.00004
2	0.10368	0.00005	0.11380	0.00007	0.11444	0.00003
3	0.12632	0.00048	0.11849	0.00075	0.14228	0.00029
4	0.11111	0.00018	0.09649	0.00029	0.14671	0.00039
5	0.09524	0.00061	0.09165	0.00085	0.13728	0.00032
6	0.07368	0.00063	0.07524	0.00062	0.12754	0.00022
7+	0.02372	0.00019	0.04709	0.00038	0.11610	0.00081

Next, I use the tree-based method on the bank dataset. I use the CART algorithm for all trees. Below is the diagram for treatment of 1 contact by the bank. Diagrams for treatments 2, 3 ,4, 5, 6, and 7+ are similar.



Below is a table that contains the calculations required to determine the average treatment effect for 1 contact using the tree-based method. Tables for the other treatments are similar.

Treatment 1		
Node	Treated Success Rate	Weighted Success Rate
Five	0.03973168	0.1193662
Seven	0.07923497	
Eight	0.3712375	
Nine	0.0788084	
Eleven	0.2251712	
Twelve	0.4219114	
Thirteen	0.2157559	

Finally, the last table includes the average treatment effect for all 7 treatments.

# of Contacts	Weighted Success Rate
1	0.1193662
2	0.110916
3	0.1443517
4	0.1411046
5	0.1392108
6	0.1123825
7+	0.1074705

Conclusion

- The Doubly Robust and Tree-Based Method seem offer the best results. The Naïve method and Inverse Propensity Weighting method give significantly lower (and likely biased) estimates.
- Theory suggests that the Doubly Robust method will have low bias and low variance. Variances for the doubly robust method are slightly lower than for the inverse propensity weighting estimates, though they are comparable.
- Inverse propensity weighting here does not result in a balance of covariates across controls, which is likely why the results are biased.
- **Future Work:**
  - More work on the tree-based method. Combine with bagging.
  - Apply to larger and more complex datasets. It would be interesting to see how these methods perform on high dimensional datasets from online advertising with millions of observations.

References

- Dudík, Miroslav, Dumitru Erhan, John Langford, and Lihong Li. "Doubly robust policy evaluation and optimization." *Statistical Science* 29, no. 4 (2014): 485-511
- Lambert, Diane, and Daryl Pregibon. "More bang for their bucks: Assessing new features for online advertisers." In *Proceedings of the 1st international workshop on Data mining and audience intelligence for advertising*, pp. 7-15. ACM, 2007.
- Moro, Sérgio, Paulo Cortez, and Paulo Rita. "A data-driven approach to predict the success of bank telemarketing." *Decision Support Systems* 62 (2014): 22-31.
- Wang, Pengyuan, Wei Sun, Dawei Yin, Jian Yang, and Yi Chang. "Robust tree-based causal inference for complex ad effectiveness analysis." In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, pp. 67-76. ACM, 2015.