GR 5243 Applied Data Science

Project 4:
Causal Inference Algorithms Evaluation

Group #4: Zhenglei Chen, Jaival Desai, Qinzhe Hu, Levi Lee, Luyao Sun, Xinyi Wei

Perform Date: 12/02/2020

Contents

A=Treatment
Y=Response
V=Variable

Low Dimension Data
475 Rows
24 Columns

High Dimension Data
2000 Rows
187 Columns

GRADUATE SCHOOL
OF ARTS AND SCIENCES

Data source: *Atlantic Causal Inference Conference (ACIC) Data Challenge*

Classification & Estimation

- Classification/Regression Trees
- $\hat{f}(x) = \sum_{m=1}^{M} c_m I\{x \in R_m\}$
- Split the space into M regions, and model the response by c_m in each region.
- The tree model choose the variable automatically to split on to achieve the best fit.
- One or both of these regions are split into two more regions, and this process is continued, until some stopping rule is applied.
- We use the Gini index for our impurity measure .

- Propensity Scores
- $e(x) = \Pr(T = 1|X = x), 0 < e(x) < 1$

- Average Treatment Effect
- $\Delta_t = E(Y_1 - Y_0) = E(Y_1) - E(Y_0)$

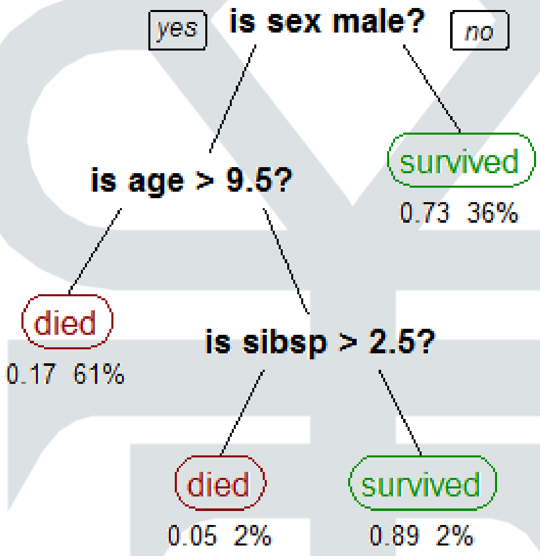
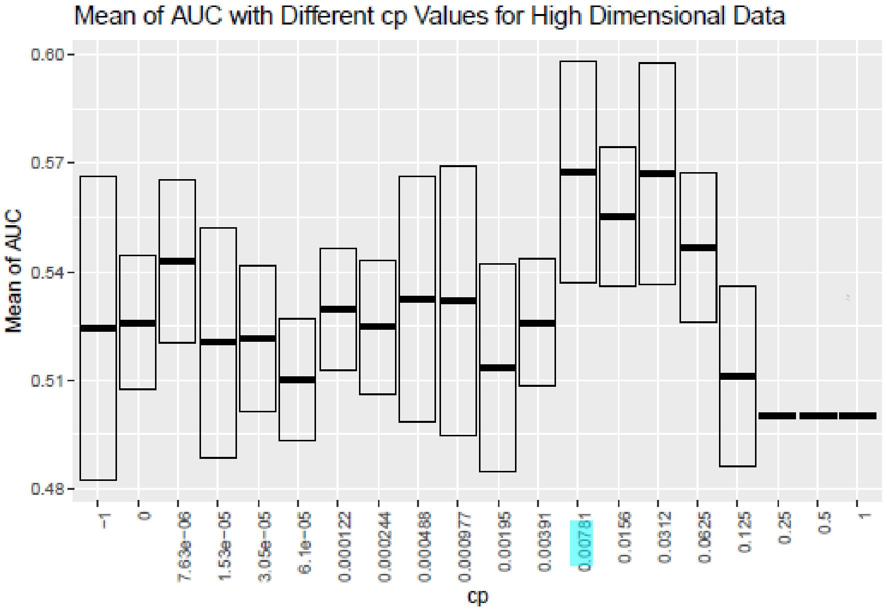
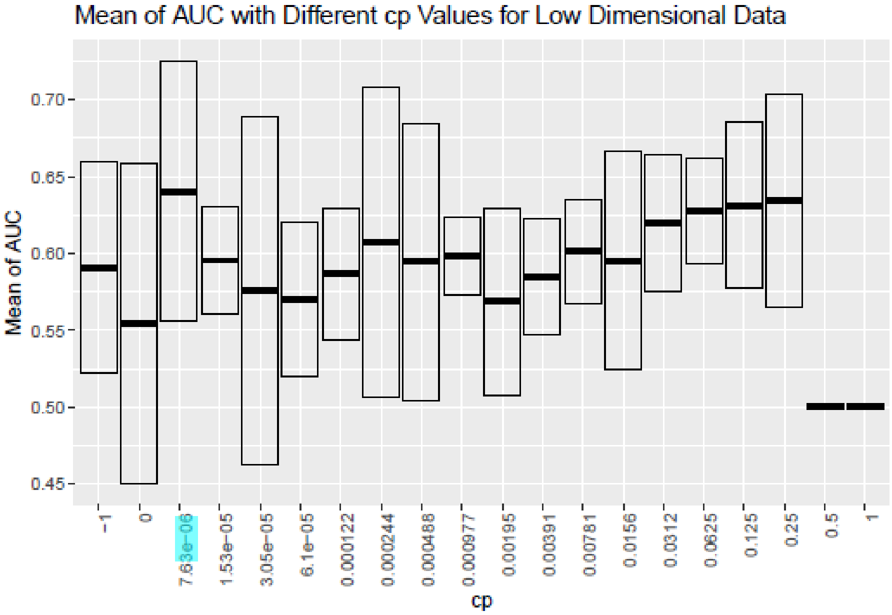


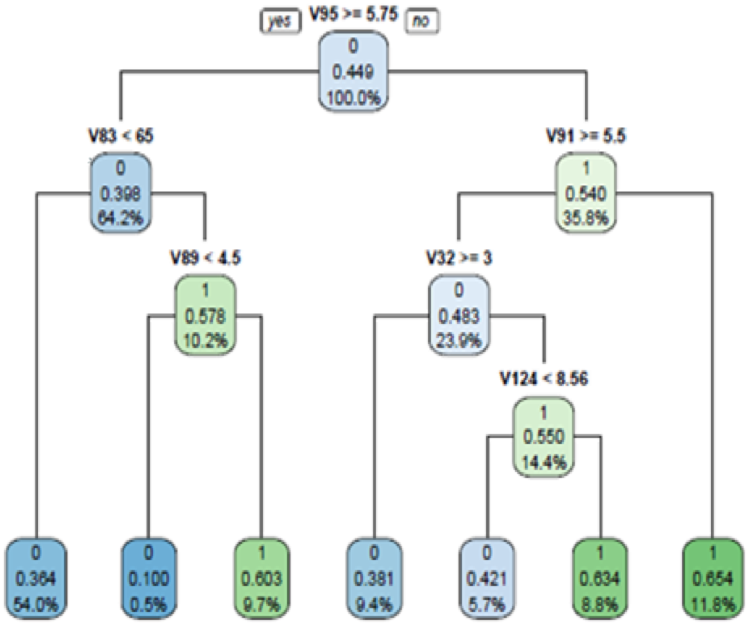
Figure: https://zh.wikipedia.org/wiki/%E5%86%B3%E7%AD%96%E6%A0%91%E5%AD%A6%E4%B9%A0

Procedure – 5 Fold Cross Validation

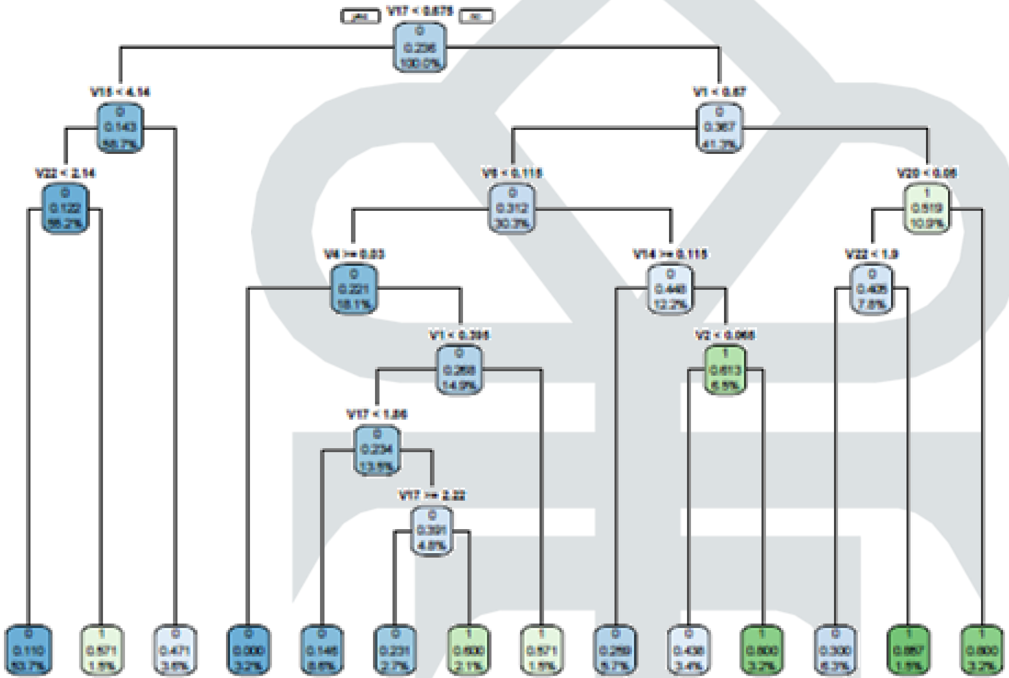




Procedure - PS Estimation



High Dimension Data



Low Dimension Data

ATE-Algorithms

- Stratification
- $\widehat{\Delta}_S = \sum_{j=1}^{K} \frac{N_j}{N} \{ \boldsymbol{N_{1j}^{-1} \sum_{i=1}^{N} T_i Y_i I(\hat{e}_i \in \widehat{Q}_j) - N_{0j}^{-1} \sum_{i=1}^{N} (1 - T_i) Y_i I(\hat{e}_i \in \widehat{Q}_j\ )} \}$
- K is the number of strata, N_j is the number of observations in stratum j. N_{1j} is the number of "treated"observations in stratum j, N_{0j} is the number of "controlled" observations in stratum j. $\hat{Q}_j$ is the sample quantile of estimated propensity scores.

- Regression Adjustment
- $Y_i = \alpha T_i + \beta PS_i + \epsilon_i$
- Regress the outcome variable Y on treatment indicator T and the estimated propensity score.

- Stratification & Regression Adjustment
- Using stratification first and then regress Y on T and PS for each stratum.
- Taking weighted average of these coefficients of T for all strata.

Results

ATE

	High Dimension Data	Low Dimension Data
True	-3.000000	2.500000
Stratification	-2.144670	2.673899
Regression Adjustment	-2.527116	3.053240
Reg Adjust. & Strati.	-2.503732	3.022839

Runtime

(NVMe SAMSUNG SSD with 16 GB RAM.)

	High Dimension Data	Low Dimension Data
PS Estimation	1.24324393	0.03291011
Stratification	0.31650686	0.12155604
Regression Adjustment	0.04108310	0.01595902
Reg Adjust. & Strati.	0.03195000	0.02493501

Conclusion

- Regression adjustment performed the best for the high dimensional data and stratification performed the best for the low dimensional data.

- Using classification/regression trees for propensity scores was not the ideal approach for either dataset.

- Different values of K, that is, the number of strata, resulted in an empty stratum in our results. Even after choosing a value of K which would present no empty strata, we saw that each stratum tend to have imbalanced classes.

- This complication may explain why the stratification plus regression adjustment method would not have performed the best.

- The results were relatively consistent among all three methods–there were no large deviations from the true value.

- All of these algorithms have very short runtime.

Reference

- Atkinson, Beth. *"Recursive Partitioning And Regression Trees."* R Documentation, DataCamp, www.rdocumentation.org/packages/rpart/versions/4.1-15/topics/rpart.
- Austin, Peter C. 2011. *"An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies."* Multivariate Behavioral Research 46 (3): 399–424.
- Chan, David & Ge, Rong & Gershony, Ori & Hesterberg, Tim & Lambert, Diane. (2010). *Evaluating online ad campaigns in a pipeline: Causal models at scale.* Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 7-16. 10.1145/1835804.1835809.
- Hastie, Trevor,, Robert Tibshirani, and J. H Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed*. New York: Springer, 2009. Print.
- D'Agostino RB Jr. *Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group*. Stat Med. 1998 Oct 15;17(19):2265-81. doi: 10.1002/(sici)1097- 0258(19981015)17:19<2265::aid-sim918>3.0.co;2-b. PMID: 9802183.
- Lunceford, Jared K, and Marie Davidian. 2004. *"Stratification and Weighting via the Propensity Score in Estimation of Causal Treatment Effects a Comparative Study."* Statistics in Medicine 23 (19): 2937–60.
- Rosenbaum PR, Rubin DB. *The central role of the propensity score in observational studies for causal efiects.*Biometrika 1983; 70:41–55.