



THE UNIVERSITY OF CHICAGO

Department of Statistics

MASTER'S THESIS PRESENTATION

ELIZABETH CLEVELAND

Department of Statistics
The University of Chicago

Study of Machine Learning Techniques for Market Research
Problems

FRIDAY, November 14, 2014, at 10:00 AM
Eckhart 117, 5734 S. University Avenue

ABSTRACT

Predicting product sales is a frequently encountered problem in market research. For this project, we explored the use of various regression techniques for the prediction of product sales. Data from Information Resources, Incorporated was used for analysis. In many market research problems, the high number of covariates makes linear models impractical and ineffective. We considered lasso regression, trees, random forests, boosted regression forests, principal components regression, and sparse principal components regression. Next, we explored the impact of outliers. Outliers are a common occurrence in market research data. To determine a model's robustness to outliers, a subset of the data was distorted to create artificial outliers. Each regression technique was analyzed to determine the effect of the imposed outliers. Finally, we considered the impact of accounting for the size of a store prior to model building.

For information about building access for persons with disabilities, please contact Laura Rigazzi at 773.702-0541 or send an email to lrigazzi@galton.uchicago.edu. If you wish to subscribe to our email list, please visit the following web site: <https://lists.uchicago.edu/web/arc/statseminars>.