Machine Learning Methods for Demand Estimation

Patrick Bajari, Denis Nekiplov, Stephen Ryan, Miaoyo Yang

American Economics Review: Papers and Proceedings (2015)

Presented by
Afrin Islam, Mursaleen Shiraj
Washington State University

February 14, 2019

Introduction

What this paper is about?

- The paper draws attention to the fact that researchers from disciplines other than Economics are also working on demand estimation.
- Discusses various methods of demand estimation that are popular in the machine learning literature
- Compares the performance of machine learning models against traditional models used in economics for demand prediction.

Traditional Economics Models

Linear Regression and Logistic Regression

- Closed form analytical solution available
- Which variables to include in the model is motivated by theory, often econometrician's choice.
- Within sample prediction power is relatively batter, out-of-sample prediction is usually not good.

Machine Learning Models

Stepwise Regression

Generates a series of nested models

1. Begin by regressing the dependent variable on an intercept and calculate the residual.
2. Find the covariate with the highest correlation with the residual.
3. Include this covariate, found in step 2, as an independent variable and regress. Get the residual.
4. Repeat from step 2 until no covariate has a sufficiently large correlation with the residuals.

Machine Learning Models

Forward stagewise regression

1. Start with regressing the dependent variable on a constant only. At this stage, it is like, coefficients on other covariates are zero. Get residuals. (variables are standardized, so mean=0)
2. Find the covariate with the highest correlation with the residual.
3. The coefficient of the covariate that has highest correlation with the residual is incremented in the direction of the correlation by a small amount (usually).
4. calculate new residual as residual from step 2 minus the weighted covariate of the covariate chosen in step 2. The weight is the coefficient estimated in step 3 times a small number.
5. Go back to step 2.

Machine Learning Models

Support Vector Machine

- Separates feature space by finding a hyperplane.
- Finds the hyperplane that maximizes the minimum distance from the data points.
- If feature space is not linearly separable, then transforms the feature space by mapping it into higher dimensional feature space so it becomes linearly separable.

Machine Learning Models

LASSO

A penalized regression method where additional regressors are penalized as the loss function (objective function) contains the sum of the absolute values of the coefficients.

Machine Learning Models

Regression Trees

- Approximate by partitioning the predictor space.
- To make prediction, uses the mean of the training sample in the region where it belongs.
- To partition, minimize the sum of the residual sum of squared in each partition.
- Use greedy algorithm to find the predictors and their respective split points.
- Often the variables and it's splitting point is chosen by considering largest drop in RSS (or highest information gain). This strategy is short sighted.
- Alternative is to grow a large tree and then prune.

Machine Learning Models

Random Forests

- Similar to regression trees
- iterate by randomly choosing a collection of predictors to make predictions.
- Iteration results in a large number of trees, hence the forest.
- Average over a large number of trees yield an unbiased prediction.

Machine Learning Models

Combined model

- Use the previously mentioned models to get estimates on the dependent variable, fitted values.
- Run a regression of the original dependent variable on these estimates.
- Use constraints to get weights between 0 and 1.

Comparing Models

TABLE 1—MODEL COMPARISON: PREDICTION ERROR

	Validation		Out-of-sample		Percent weight
	RMSE	SE	RMSE	SE	
Linear	1.169	0.022	1.193	0.020	6.62
Stepwise	0.983	0.012	1.004	0.011	12.13
Forward stagewise	0.988	0.013	1.003	0.012	0.00
LASSO	1.178	0.017	1.222	0.012	0.00
Random forest	0.943	0.017	0.965	0.015	65.56
SVM	1.046	0.024	1.068	0.018	15.69
Bagging	1.355	0.030	1.321	0.025	0.00
Logit	1.190	0.020	1.234	0.018	0.00
Combined	0.924		0.946		100.00

Concluding Remarks

- Words of caution from the authors . . .

  > " we do not claim that these models dominate all methods proposed in the voluminous demand estimation literature. Rather, we claim that as compared to common methods an applied econometrician might use in off the shelf statistical software, these methods are considerably more accurate."

- These methods are concerned with prediction only, not on structural parameter estimation.

Thank You