Study on tree-based methods.

MATH 6380 project 2

Chenyang, DONG Tsz Cheung, LO Jiacheng, XIA

April 23, 2017

Outline

Table of Contents

Studying tree based methods...

Why did we choose tree-based methods?

Studying tree based methods...

Why did we choose tree-based methods?

- We went through several methods.

Studying tree based methods...

Why did we choose tree-based methods?

- We went through several methods.
- Tree-based methods are straight-forward and easy to implement.

Studying tree based methods...

Why did we choose tree-based methods?

- We went through several methods.
- Tree-based methods are straight-forward and easy to implement.
- There are yet many improvement methods.

Studying tree based methods...

Why did we choose tree-based methods?

- We went through several methods.
- Tree-based methods are straight-forward and easy to implement.
- There are yet many improvement methods.
- Studied the method on 3 datasets.

Table of Contents

The dataset

This dataset and the preprocessing are the same as project 1.

The dataset

This dataset and the preprocessing are the same as project 1.

Goal for this dataset

Do straightforward analysis and compare with Lasso (PJ 1).

What we found

In terms of MSE, simple regression tree(0.11) slightly worse than Lasso(0.06); bagging, random forest and boosting even better(0.04, 0.02).

Visualize the results

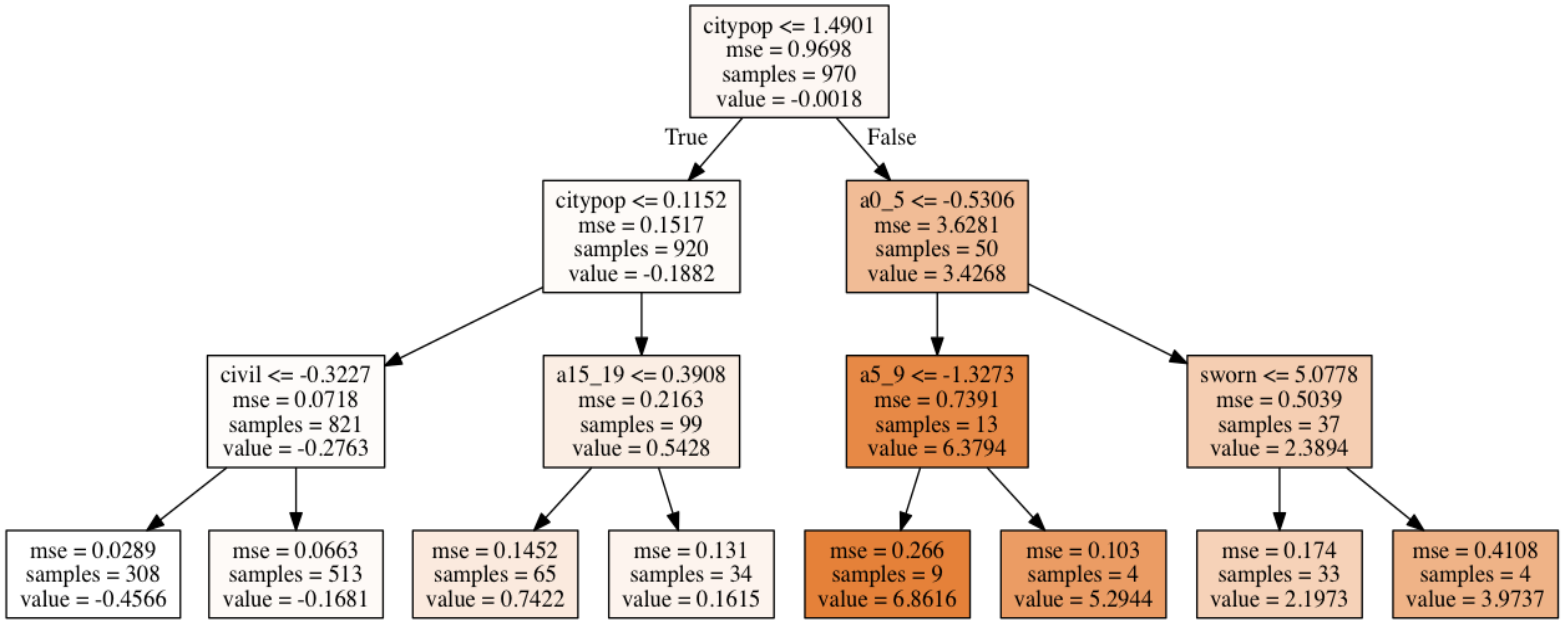

Figure: Regression tree on crime data

Boosting and random forests

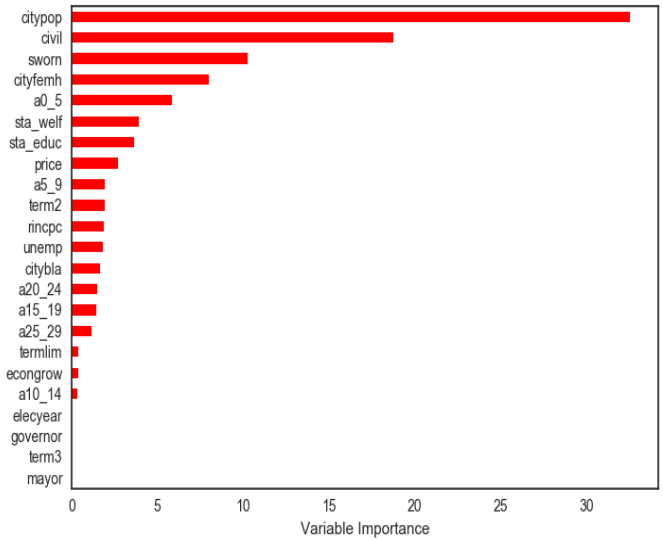

Figure: Importance from boosting

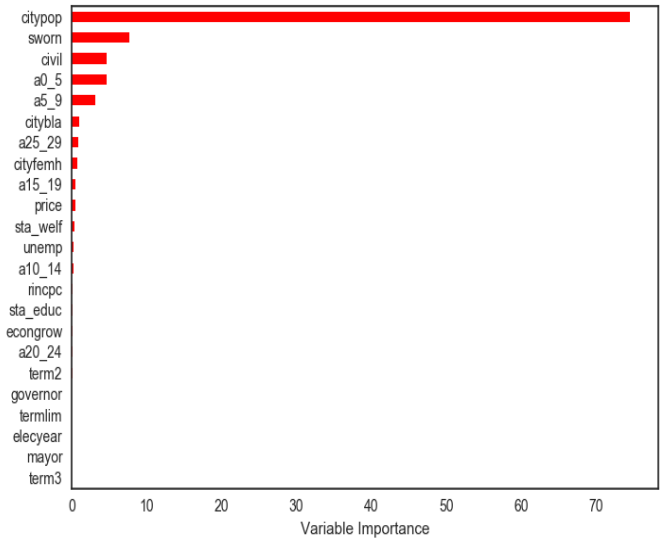

Figure: Importance from random forest

Table of Contents

Table of Contents

Models and Kaggle Results

Dummy1

Models and Kaggle Results

Dummy1 Dummy2

Variable Importance

Dummy

Compared with p-value Selection

Dummy

Some Conclusions

We came up with some conclusions (inferences):

Some Conclusions

We came up with some conclusions (inferences):

1. dummy1

Some Conclusions

We came up with some conclusions (inferences):

1. dummy1
2. dummy2

Table of Contents