

Understanding Theft in America on the County Level

Mason Shihab and Elinor Chu

12/19/21

Contents

1	Top 10 Counties	1
2	Ridge Coefficients	2
3	Boosting Importance	2
4	Model Evaluation	3

1 Top 10 Counties

```
read_tsv("../results/top-10-counties-data.tsv") %>%
  kable(format = "latex", row.names = NA,
        booktabs = TRUE,
        digits = 5,
        col.names = c("State", "County", "Theft Rate"),
        caption = "This a table showing the top 10 counties with the highest theft rate.") %>%
  kable_styling(position = "center") %>%
  kable_styling(latex_options = "HOLD_position")
```

```
## Rows: 10 Columns: 3
```

```
## -- Column specification -----
## Delimiter: "\t"
## chr (2): state, county
## dbl (1): theftrate
```

```
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

Table 2: Standardized coefficients for features in the Ridge model based on the one-standard-error rule.

Feature	Coefficient
housing_density	0
pop_density	0
pertrump	0
unemp_bens_possible	0
police_funding_score	0
gini	0
Marriedcouplefamily	0
no_health_ins	0
pct_child_in_pov	0
svi_overall	0

Table 1: This a table showing the top 10 counties with the highest theft rate.

State	County	Theft Rate
Indiana	Hamilton	0.10498
Nebraska	Sarpy	0.08697
New York	New York	0.07927
Colorado	Jefferson	0.07638
Colorado	Denver	0.04860
Mississippi	Tunica	0.04465
California	San Francisco	0.04427
Missouri	Greene	0.04167
New Mexico	Bernalillo	0.04156
West Virginia	Wayne	0.04136

2 Ridge Coefficients

3 Boosting Importance

```
read_tsv("../results/gbm_opt_rel_imp.tsv") %>%
  kable(format = "latex", row.names = NA,
        booktabs = TRUE,
        digits = 5,
        col.names = c("Variable", "Relative influence"),
        caption = "These are the first ten rows of the relative influence
        table for the optimal boosting model above.") %>%
  kable_styling(position = "center") %>%
  kable_styling(latex_options = "HOLD_position")
```

```
## Rows: 12 Columns: 2
```

```
## -- Column specification -----
```

Table 4: Root-mean-squared prediction errors.

Model	Test_RMSE
Boosting	0.0047690
Intercept-only	0.0048906
Ridge	0.0049133
Lasso	0.0049133
Elastic_Net	0.0049133
Random_Forest	0.0049465

```
## Delimiter: "\t"
## chr (1): var
## dbl (1): rel.inf

##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

Table 3: These are the first ten rows of the relative influence table for the optimal boosting model above.

Variable	Relative influence
unemp_bens_possible	10.4522
PctEmpFIRE	9.5739
bachplus	8.8310
poor_fair_health	8.3525
pertrump	7.1198
dis35to64	6.3297
PopChangeRate1819	6.2759
PerCapitaInc	5.0603
housing_density	4.7507
no_health_ins	3.5819
police_accountability_score	3.4705
inundergrad	3.0355

4 Model Evaluation