The Conference Name Here

Your Descriptive Title Here

FirstName1 LastName1, FirstName2 LastName2, FirstName3 LastName3
Department of Mathematical Sciences
Appalachian State University, Boone, North Carolina

Overview

- On May 8, 2012, North Carolina voters approved Amendment One. This poster examines four different models used to predict North Carolina county voting behavior.
- To ensure accurate predictive power for future observations, the data are split into a training set (80%) and a test set (20%).
- Root mean squared error of the test set is used as a measure of model adequacy.
- All computations and graphs are created with the open source software R [4].
- Another bullet here
- Yet a fifth bullet here
- A sixth bullet if you want

K-Fold Cross-Validation

- Cross validation is the simplest and most widely used method for estimating prediction error [2]. This method directly estimates the expected extra-sample error, $Err = E[L(Y, \hat{f}(X))]$. In this work, the loss function, L, is the square root of the average squared error loss.
- The data in this project is split into $K = 10$ equal sized parts. The cross-validation estimate of the prediction error is

$$CV(\hat{f}) = \frac{1}{N}\sum_{i=1}^{N} L(y_i, \hat{f}^{-K(i)}(x_i)),$$

where $\hat{f}^{-K}(x)$ denotes the fitted function with the K^{th} part of the data removed.

Basic Models Used

I. Least Squares Regression
Note: Variables are described in the Variable Table handout.
   a. Model from [1] applied to Training data (mod1A) — Percent voting for Amendment One is modeled using the predictors pct18.24, medinc, pctb, mccain08, evanrate, pctrural, and pctba.
   b. Our OLS model applied to Training data (mod1B) — Percent voting for Amendment One is modeled using the predictors obama08, pctrural, pctw, pctd, pctb, log(pct18.24), log(pctcolenrol), pctfm, log(pctfd), pctown, medinc, medinc^2, evanrate, evanrate^2, pctfb, pctfb^2, log(pctstud), and log(colden).

II. Cross validated, $K = 10$, and pruned, $n_{\text{leaves}} = 4$, regression tree [2] (mod2)

III. Random Forest built from, $n_{\text{trees}} = 5000$, [3] (mod3)

Resources

1. Our incredible web page
2. Our incredible WOODCARB3R package

Overview

- On May 8, 2012, North Carolina voters approved Amendment One. This poster examines four different models used to predict North Carolina county voting behavior.
- To ensure accurate predictive power for future observations, the data are split into a training set (80%) and a test set (20%).
- Root mean squared error of the test set is used as a measure of model adequacy.
- All computations and graphs are created with the open source software R [4].

K-Fold Cross-Validation

- Cross validation is the simplest and most widely used method for estimating prediction error [2]. This method directly estimates the expected extra-sample error, $Err = E[L(Y, \hat{f}(X))]$. In this work, the loss function, L, is the square root of the average squared error loss.
- The data in this project is split into $K = 10$ equal sized parts. The cross-validation estimate of the prediction error is

$$CV(\hat{f}) = \frac{1}{N}\sum_{i=1}^{N} L(y_i, \hat{f}^{-K(i)}(x_i)),$$

where $\hat{f}^{-K}(x)$ denotes the fitted function with the K^{th} part of the data removed.

Basic Models Used

I. Least Squares Regression
Note: Variables are described in the Variable Table handout.
   a. Model from [1] applied to Training data (mod1A) — Percent voting for Amendment One is modeled using the predictors pct18.24, medinc, pctb, mccain08, evanrate, pctrural, and pctba.
   b. Our OLS model applied to Training data (mod1B) — Percent voting for Amendment One is modeled using the predictors obama08, pctrural, pctw, pctd, pctb, log(pct18.24), log(pctcolenrol), pctfm, log(pctfd), pctown, medinc, medinc^2, evanrate, evanrate^2, pctfb, pctfb^2, log(pctstud), and log(colden).

II. Cross validated, $K = 10$, and pruned, $n_{\text{leaves}} = 4$, regression tree [2] (mod2)

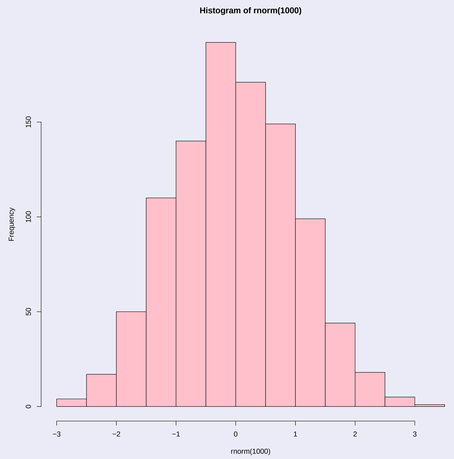
III. Random Forest built from, $n_{\text{trees}} = 5000$, [3] (mod3)

Incredible R Code

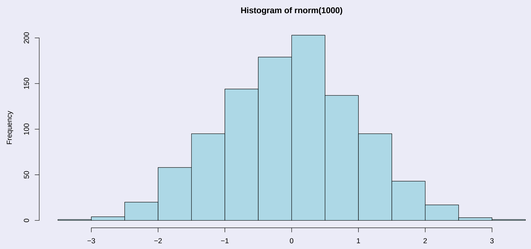
```
N <- 10000
x <- rnorm(100, 0, 1)
results <- numeric(N)
for(i in 1:N){
  bs <- sample(x, size = 100, replace = TRUE)
  results[i] <- mean(bs)
}
mean(results)
```

```
[1] -0.08555566
```

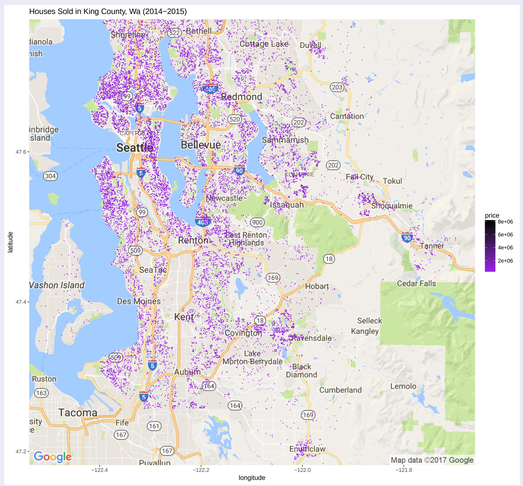
A Pink Histogram



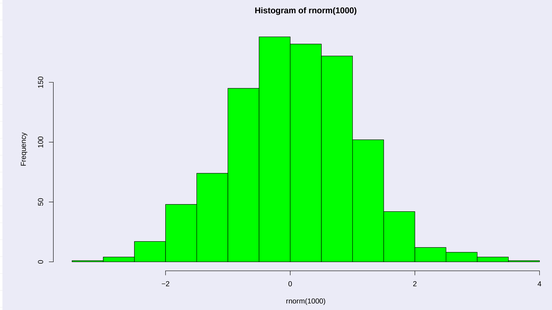
Some Title



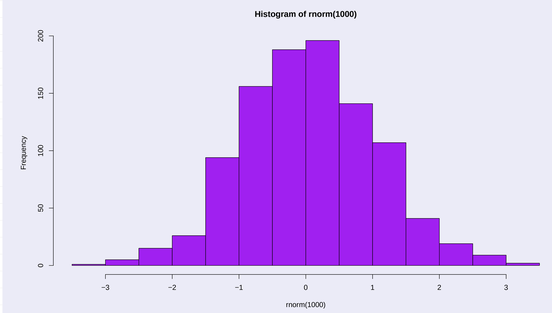
King County Real Estate



Random Forest Variable Importance



Prediction Errors



Further Directions

- Use the models developed in this poster to predict county votes for states that have had similar marriage amendments such as South Carolina, Wisconsin, South Dakota, Florida, Idaho, Alabama, Utah, Michigan, Texas, Arkansas, Louisiana, Kansas, Kentucky, Ohio, and Nebraska.
- Use ensemble methods (combining multiple models) for better prediction.
- Make our local maps available via the internet using the shiny server.

References

[1] Elizabeth Davison and Jessica Eatman.
An ecological examination of north carolina's amendment one vote to ban same sex marriage.
2013.

[2] Trevor Hastie, Robert Tibshirani, and Jerome Friedman.
The elements of statistical learning, volume 2.
Springer, 2009.

[3] Andy Liaw and Matthew Wiener.
Classification and regression by randomForest.
R news, 2(3):18–22, 2002.

[4] R Core Team.
R: A Language and Environment for Statistical Computing.
R Foundation for Statistical Computing, Vienna, Austria, 2016.

Faculty Advisors: Eric Marland & Alan Arnholt

https://github.com/alanarnholt/USFS/