Reproducible Research and Teaching in the Cloud

Gordon David Mosher and Thomas Girke
University of California, Riverside - Departments of Statistics and Bioinformatics

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.
- All computations and graphs are created with the open source software `R` [**?**].

K-Fold Cross-Validation

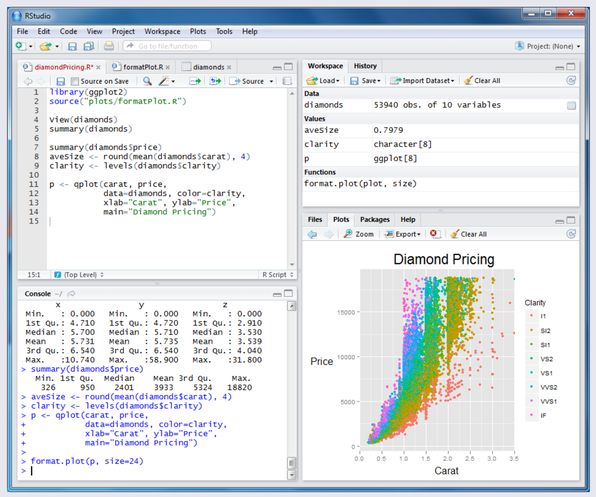
[**RStudio takes your markdown**]

- **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 Pink Histogram

A Pink Histogram

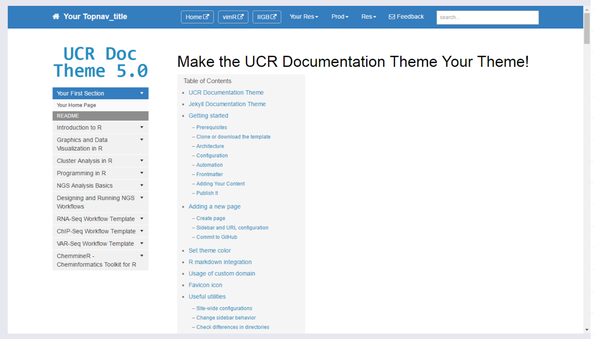
```
  Sepal.Length Sepal.Width
1          5.1         3.5
2          4.9         3.0
3          4.7         3.2
4          4.6         3.1
5          5.0         3.6
  Petal.Length Petal.Width Species
1          1.4         0.2  setosa
2          1.4         0.2  setosa
3          1.3         0.2  setosa
4          1.5         0.2  setosa
5          1.4         0.2  setosa
  Sepal.Length    Sepal.Width
 Min.   :4.300   Min.   :2.000
 1st Qu.:5.100   1st Qu.:2.800
 Median :5.800   Median :3.000
 Mean   :5.843   Mean   :3.057
 3rd Qu.:6.400   3rd Qu.:3.300
```

References

[1] R Core Team.
R: A Language and Environment for Statistical Computing.
R Foundation for Statistical Computing, Vienna, Austria, 2015.

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] R Core Team.
R: A Language and Environment for Statistical Computing.
R Foundation for Statistical Computing, Vienna, Austria, 2015.

WWW: `https://bioconductor.org/packages/bioassayR` Mail: `tbackman@ucr.edu`