STOR 320 Modeling XI

Lecture 34

Yao Li
Department of Statistics and Operations Research
UNC Chapel Hill

Shrinkage Estimation

- Three Popular Methods
  - Download R Package `> library(glmnet)`
  - Penalized SSE

$$PSSE = SSE + \lambda[(1-\alpha)\sum_{i=1}^{p}\beta_i^2 + \alpha\sum_{i=1}^{p}|\beta_i|]$$

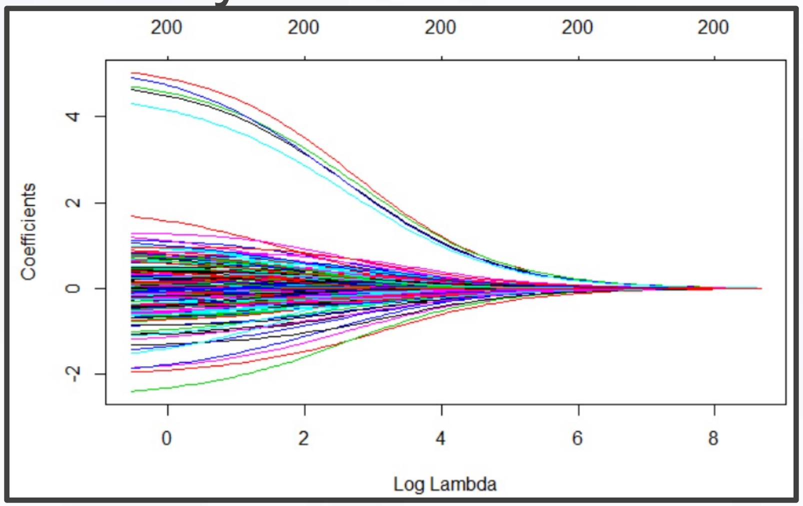
  - Variations
    - Ridge (1970): $\lambda = 1 \,\&\, \alpha = 0$
    - Lasso (1996): $\lambda = 1 \,\&\, \alpha = 1$
    - Elastic Net (2005)

$$\lambda = 1 \;\&\; 0 < \alpha < 1$$

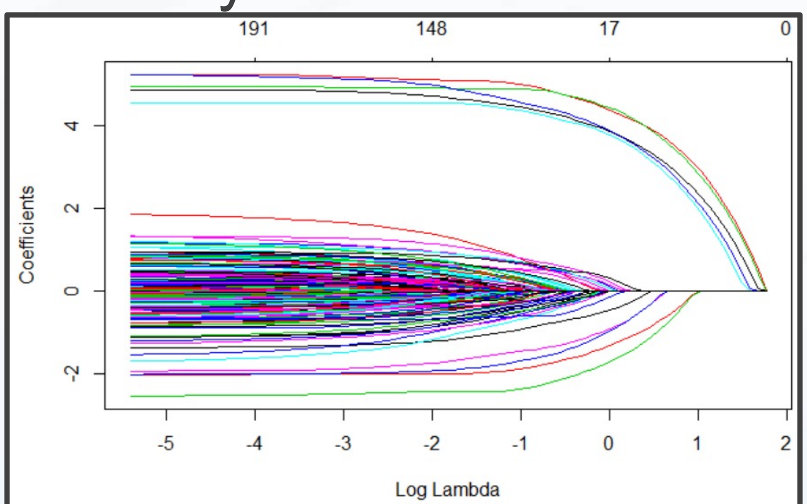
  - Notice When
    - $\lambda = 0$ PSSE=SSE
    - As λ Gets Bigger, the Coefficients Approach 0

Ridge vs. Lasso vs. Elastic Net

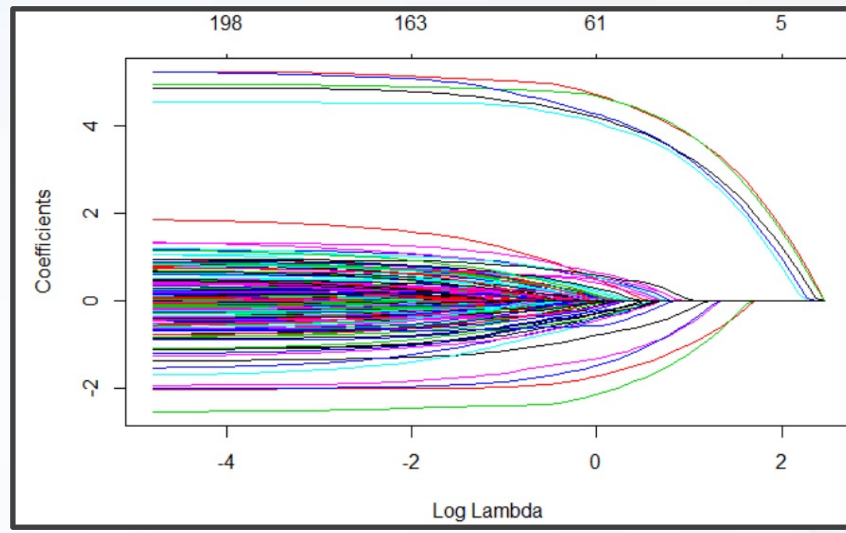
- Ridge Penalty



- Lasso Penalty



- Elastic Net Penalty



Next Steps

- Tuning Parameters
  - Use Cross-Validation to Choose Tuning Parameters $\lambda \,\&\, \alpha$
  - Constraints
    - $\lambda > 0$
    - $0 \leq \alpha \leq 1$
  - Best Approach:
    - Divide Data Into Train & Test
    - Loop Over a Vector of Alpha
    - Find Best Lambda for Each Alpha Considered Using CV in Train
    - For Each Alpha and Best Lambda, Predict on Test and Select Alpha and Lambda that Minimize MSE

Part 2-Chunk 4

- Run Chunk 4
  - Illustration of 10 Fold CV
  - Finding Best Combination of Alpha and Lambda

alpha	lambda	MSE
0.0	17.282127	176.3021
0.1	7.837234	146.4758
0.2	5.180181	139.9872
0.3	3.453454	133.7793
0.4	2.590091	130.7873
0.5	2.495819	132.6983
0.6	1.895081	129.1495
0.7	1.624355	128.1601
0.8	1.559887	129.2083
0.9	1.386566	128.5799
1.0	1.247909	128.0857

Best: $\alpha = 1$ & $\lambda = 1.25$

Part 2-Chunk 5

- Run Chunk 5
  - The Top 4 Models

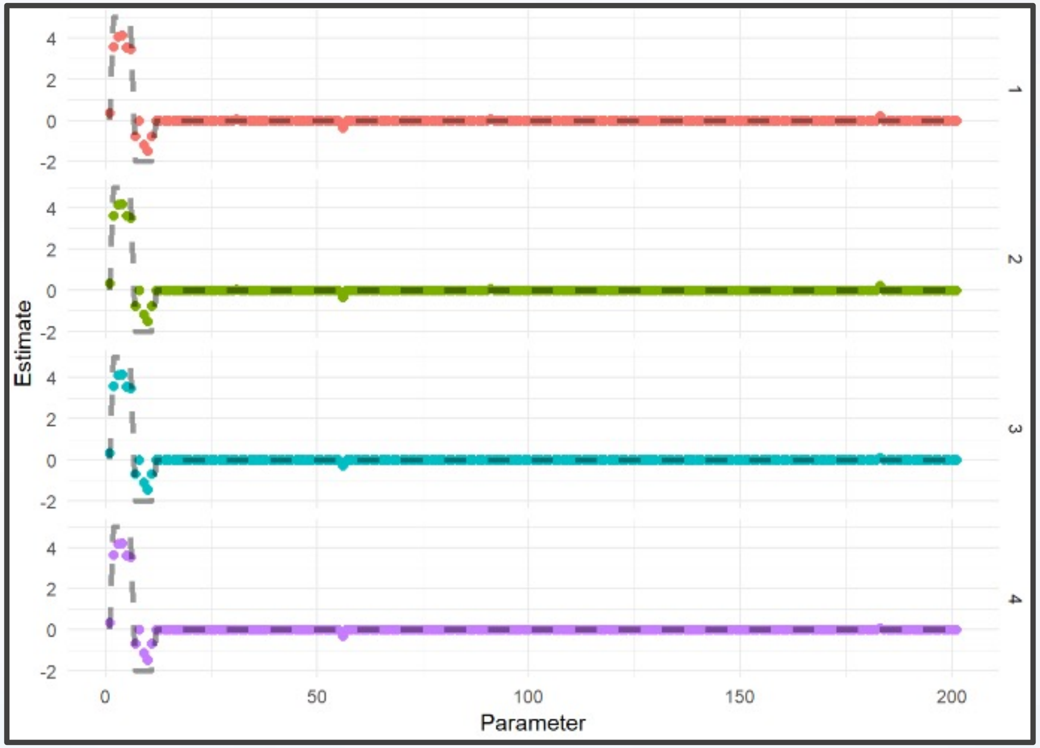
```
##    alpha   lambda      MSE
## 1    0.6 1.895081 129.1495
## 2    0.7 1.624355 128.1601
## 3    0.9 1.386566 128.5799
## 4    1.0 1.247909 128.0857
```

  - Question: How Different Are These Models?

  - For Each Alpha & Lambda,
    - Get Final Coefficients
    - Compare Across Models
    - Compare to True Values

Part 2-Chunk 5

- Chunk 5 (Continued)
  - Visualizing Top Four
    - Points Show Estimates
    - Dashed Line Shows Truth



Part 3: More Application

- Built-In Data `> mpg`
  - n=234
  - Focus is on Modeling Hwy MPG
  - Subset Data to Include Only Wanted Covariates

year <int>	displ <dbl>	cyl <int>	drv <chr>	cty <int>	hwy <int>	fl <chr>	class <chr>
1999	1.8	4	f	18	29	p	compact
1999	1.8	4	f	21	29	p	compact
2008	2.0	4	f	20	31	p	compact
2008	2.0	4	f	21	30	p	compact
1999	2.8	6	f	16	26	p	compact
1999	2.8	6	f	18	26	p	compact

  - There are p=7 Covariates
  - Difficulty
    - Fitting all Combinations
    - Considering All 2-Way Interaction Terms

Part 3-Chunk 1

- Run Chunk 1
  - Creating Model Matrix
    - Up to 2-Way Interactions
    - Now, p=114
  - Model Selection is Difficult
  - Dividing Data into Train & Test is Not Advised (n=234)
- Run Chunk 2
  - Only a Few Options

alpha <dbl>	lambda <dbl>	CV.Error <dbl>
0.00	1.44063441	1.722966
0.25	0.55006214	1.620769
0.50	0.18956825	1.488094
0.75	0.10492193	1.456773
1.00	0.04942052	1.411025

Lowest Estimation of Prediction Error

Part 3-Chunk 2

- Chunk 2 (Continued)
  - Understanding cv.glmnet Object
    - $lambda = Contains Vector of Lambda Auto-Generated
    - $cvm = Cross Validated Estimate of Error for Each Lambda in $lambda
    - $lambda.min = The Lambda that Leads to Smallest CV Measure of Error
    - $lambda.1se = The Largest Value of Lambda Such That Error is Within 1 SD of the Error Using $lambda.min

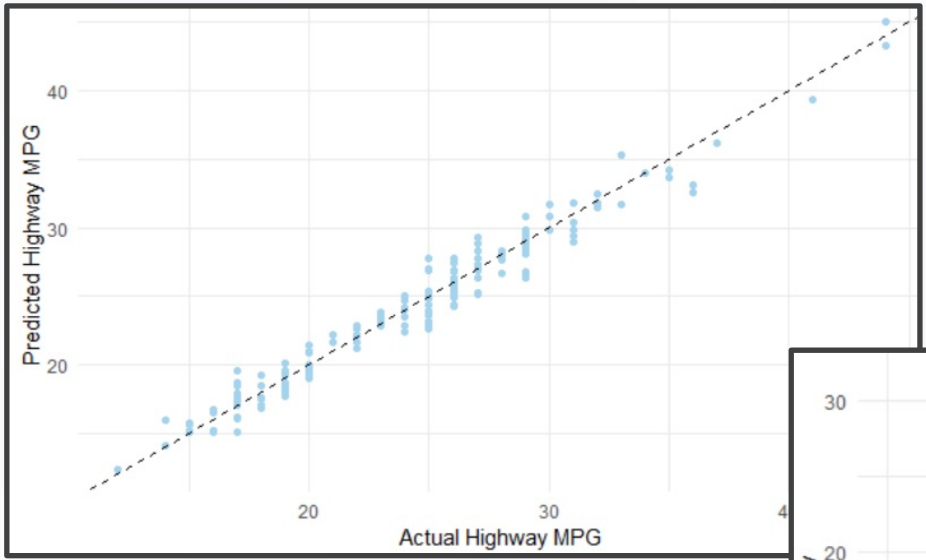
Part 3-Chunk 3

- Run Chunk 3
  - Next
    - Use Best Alpha and Lambda
    - Observe the Non-Zero Coefficients
    - Plot Predictions and Errors
  - Table of Non-Zero Coefficients
    - Before p=114
    - Now p=28

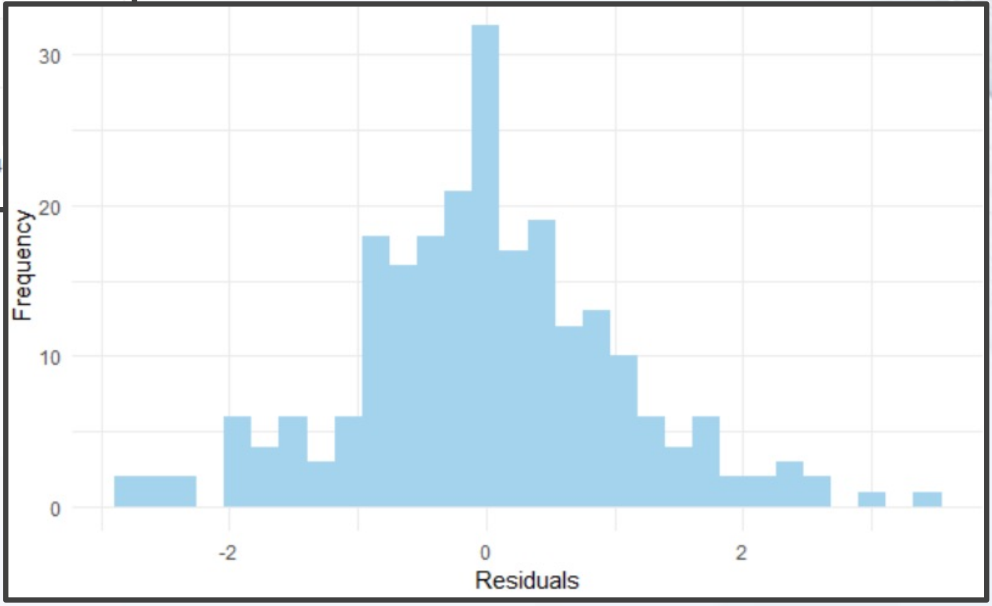
```
## # A tibble: 29 x 2
##    Parameter              Estimate
##    <chr>                     <dbl>
##  1 Int                   -123.
##  2 year                     0.0660
##  3 cty                      0.799
##  4 fle                     -1.37
##  5 flr                     -0.0629
##  6 classpickup             -0.104
##  7 classsuv                -1.37
##  8 year:cyl                -0.0000392
##  9 year:drvf                0.0000955
## 10 year:cty                 0.0000565
## 11 year:classmidsize        0.0000259
## 12 year:classpickup        -0.000659
## 13 displ:drvr               0.127
## 14 displ:classmidsize       0.0317
## 15 displ:classsuv          -0.178
## 16 cyl:fle                 -0.143
## 17 cyl:flr                 -0.0973
## 18 cyl:classcompact         0.0462
## 19 cyl:classsuv            -0.0262
## 20 drvf:cty                 0.0466
## 21 drvr:cty                 0.0282
## 22 drvf:fld                 2.54
## 23 drvr:classsubcompact    -0.0754
## 24 cty:classminivan        -0.0574
## 25 cty:classpickup         -0.106
## 26 flr:classmidsize         0.488
## 27 flp:classsubcompact     -1.42
## 28 fld:classsuv            -0.552
## 29 flp:classsuv            -0.431
```

Part 3-Chunk 3

- Comparing Predict and Actual



- Distribution of Residuals



Part 3:Example 2

- Ecdat Data `> Participation`
  - Labor Market Participation of Married Women in Switzerland
  - Data From 1981
    - 872 Married Women
    - Variables
      - Participation (Binary)
      - Non-Labor Income (log transformed)
      - Age (Scaled by 10)
      - Education (Years)
      - # of Young Children
      - # of Older Children
      - Foreigner (Binary)

Part 3-Chunk 4

- Run Chunk 4
  - Observe the Data

	lfp	lnnlinc	age	educ	nyc	noc	foreign
1	no	10.78750	3.0	8	1	1	no
2	yes	10.52425	4.5	8	0	1	no
3	no	10.96858	4.6	9	0	0	no
4	no	11.10500	3.1	11	2	0	no
5	no	11.10847	4.4	12	0	2	no
6	yes	11.02825	4.2	12	0	1	no

  - We Would Like to Build a Model to Predict Labor Involvement

  - Method: Logistic Regression

Part 3-Chunk 5

- Run Chunk 5
  - Only a Few Options

alpha <dbl>	lambda <dbl>	CV.Error <dbl>
0.00	0.18805333	0.3405963
0.25	0.06542368	0.3291284
0.50	0.04745928	0.3348624
0.75	0.03472433	0.3348624
1.00	0.02604325	0.3348624

Lowest Estimation of Prediction Error

  - Notice Using Binomial Family
  - What is the Purpose of the Following?

```
type.measure="class"
```

Part 3-Chunk 6

- Run Chunk 6
  - Only Considering Best Choices

  - Observe the Coefficients
    - Useful Variables?
    - Useless Variables?

  - Observe the Confusion Matrix
    - Misspecification Error Matches What We Saw
    - $0.329 = \frac{78+209}{393+78+209+192}$

  - Write Code That Counts
    - # of Labor Participants
    - # of Predicted Participants