Homework 8

Jeff Tilton

10/13/2018

Question 11.1

Using the crime data set uscrime.txt from Questions 8.2, 9.1, and 10.1, build a regression model using: 1. Stepwise regression 2. Lasso 3. Elastic net For Parts 2 and 3, remember to scale the data first – otherwise, the regression coefficients will be on different scales and the constraint won't have the desired effect.

Goals

1. Use 3 regression techniques (Stepwise, LASSO, and Elastic Net) on the US crime dataset
2. Choose a model

Stepwise Regression

Is a classic form of variable regression. There are many forms, but one common technique is to add a new factor to the model and after adding each additional factor previously added factors are eliminated that no longer appear to be contributing to the model based on the p-value, R^2 or some other type of quality indicator.

Lasso Regression

LASSO regression uses the standard regression equation subject to a constraint on the coefficients as seen below.

$$min \sum_{i=1}^{n} (y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + ... + a_j x_{ji})^2$$
$$s.t. \sum_{i=1}^{j} |a_i| \leq \tau$$

where τ is picked at the user's discretion. It is important to scale the data before performing LASSO because the units will artificially affect how big the coefficients need to be.

Elastic Net Regression

Constrains a combination of absolute value of coefficients and their squares

$$min \sum_{i=1}^{n} (y_i - (a_0 + a_1 x_{1i} + a_2 x_{2i} + ... + a_j x_{ji})^2$$
$$s.t. \sum_{i=1}^{j} |a_i| + (1 - \lambda) \sum_{i=1}^{j} a_i^2 \leq \tau$$

Choose the best model

Method

1. Get model attributes using various regression techniques
2. Create new models with the chosen predictors
3. Use 10-fold cross-validation to obtain mean squared error (mse)
4. Choose model with lowest mse

Final Models

Step Wise Regression

```
##
## Call:
## lm(formula = Crime ~ ., data = step_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -444.70 -111.07    3.03  122.15  483.30
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6426.10    1194.61  -5.379 4.04e-06 ***
## M              93.32      33.50   2.786  0.00828 **
## Ed            180.12      52.75   3.414  0.00153 **
## Po1           102.65      15.52   6.613 8.26e-08 ***
## M.F            22.34      13.60   1.642  0.10874
## U1          -6086.63    3339.27  -1.823  0.07622 .
## U2            187.35      72.48   2.585  0.01371 *
## Ineq           61.33      13.96   4.394 8.63e-05 ***
## Prob        -3796.03    1490.65  -2.547  0.01505 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 195.5 on 38 degrees of freedom
## Multiple R-squared:  0.7888, Adjusted R-squared:  0.7444
## F-statistic: 17.74 on 8 and 38 DF,  p-value: 1.159e-10
```

Lasso Regression

```
##
## Call:
## lm(formula = Crime ~ ., data = lasso_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -432.15 -120.60    4.08  127.39  555.94
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5616.1314  1338.9944  -4.194 0.000158 ***
## M              98.4869    35.7175   2.757 0.008906 **
```

```
## Ed           169.8037    54.9734   3.089 0.003746 **
## Po1          116.6600    14.1253   8.259 5.23e-10 ***
## M.F           10.4164    13.6995   0.760 0.451741
## U2            79.2787    43.4032   1.827 0.075629 .
## Ineq          65.7785    14.3900   4.571 5.02e-05 ***
## Prob       -3947.1524  1932.3554  -2.043 0.048068 *
## Time          -0.4702     6.0712  -0.077 0.938679
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 203.9 on 38 degrees of freedom
## Multiple R-squared:  0.7704, Adjusted R-squared:  0.7221
## F-statistic: 15.94 on 8 and 38 DF,  p-value: 5.317e-10
```

Elastic Net Regression

```
##
## Call:
## lm(formula = Crime ~ ., data = e_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -457.95  -92.49    7.46  125.79  491.68
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6304.581   1407.594  -4.479 7.65e-05 ***
## M              86.841     39.632   2.191 0.035183 *
## Ed            186.888     58.158   3.213 0.002815 **
## Po1            98.833     18.284   5.405 4.70e-06 ***
## LF           -389.665   1190.413  -0.327 0.745364
## M.F            25.094     17.776   1.412 0.166873
## NW              2.085      5.422   0.385 0.702932
## U1          -6512.710   3638.847  -1.790 0.082146 .
## U2            186.887     75.957   2.460 0.018957 *
## Ineq           59.754     16.001   3.735 0.000668 ***
## Prob        -4419.448   2085.472  -2.119 0.041245 *
## Time           -1.567      6.239  -0.251 0.803152
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 202.9 on 35 degrees of freedom
## Multiple R-squared:  0.7906, Adjusted R-squared:  0.7247
## F-statistic: 12.01 on 11 and 35 DF,  p-value: 6.983e-09
```

Discussion

	Stepwise	LASSO	Elastic Net
Cross-Validated MSE	4.55e+05	4.13e+05	2.70e+05
Final Model R^2	0.79	0.77	0.79
Final Model Adjusted R^2	0.74	0.72	0.72

Elastic Net had the lowest mse as seen in the table above and tied with LASSO for the lowest adjusted R^2. T

Stepwise regression and LASSO both chose 8 predictors, but stepwise chose unemployment rate of urban males 35–39 (U2) and LASSO chose average time in months served by offenders in state prisons before their first release (Time), elastic ended up with the most predictors (11).