STA 9891 -HW 4

Fall 2021 - Prof. Rad

Erik Carrion

11/7/2021

Contents

Part I & II - Top 5 Positive & Negative Words

In order to determine the Top 5 Positive and Negative words for Elastic Net, Lasso, and Ridge we first had to cross validate our tuning parameter to find an optimal value of Lambda for each of the three models. Doing the analysis proved to be computationally intensive, requiring 4.405 hours to find a solution.

The top 5 Positive & Negative words for Elastic Net, Lasso, and Ridge Penalized Logistic Regression are below. The first 3 columns are the positive words and the last 3 columns are the negative words. Note Elastic Net and Lasso are in 100% agreement.

```
##   Elnet.Positive Lasso.Positive Ridge.Positive Elnet.Negative Lasso.Negative Ridge.Negative
## 1              7              7            gem         ugghhh         ugghhh         ugghhh
## 2     refreshing     refreshing       captures     ravenously     ravenously       microbes
## 3    wonderfully    wonderfully           noir        paris's        paris's   legislatures
## 4       captures       captures    wonderfully        'always        'always        faulted
## 5           noir           noir     refreshing        faulted        faulted        paris's
```

Part III - ROC Curves & AUC

The AUC for Elastic Net, Lasso, and Ridge are below. We can see that all three models perform roughly equally well in terms of AUC. Numerically, Ridge performs the best in Training and Testing, but its average margin over the others is only 0.00112.

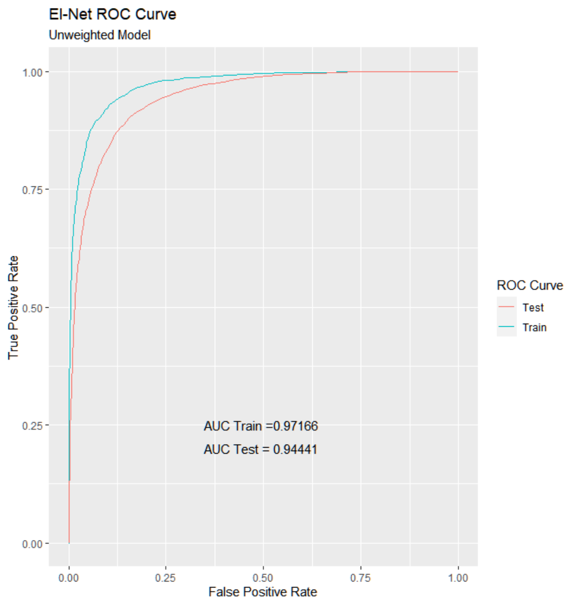
```
##        Method Train.AUC  Test.AUC Difference
## 1 Elastic Net 0.9716640 0.9444121 0.02725198
## 2       Lasso 0.9717255 0.9439869 0.02773866
## 3       Ridge 0.9729209 0.9452150 0.02770596
```

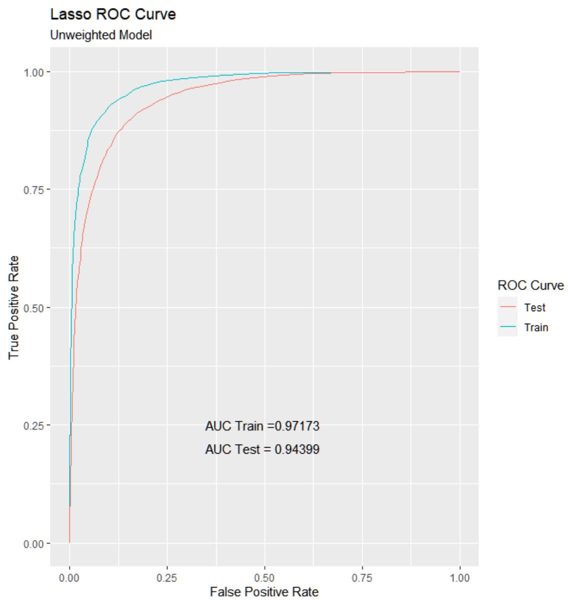
The ROC curves reflect what we see above in terms of each model's AUC.

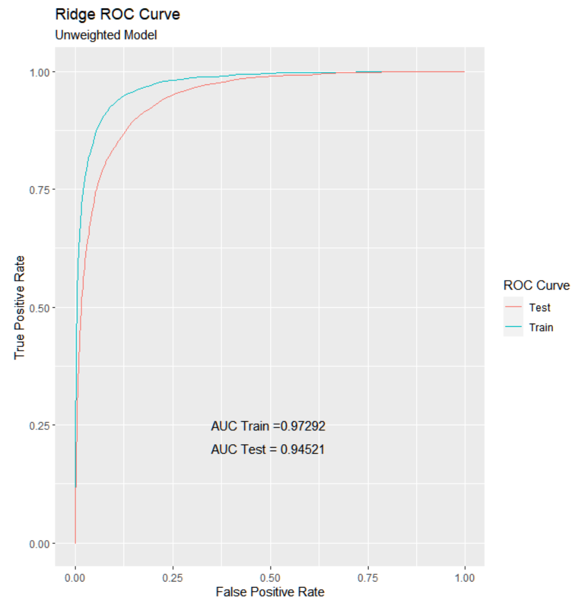
We can attribute Ridge's better AUC to the fact it uses 100% of the feature space. Elastic Net uses both the L1 and L2 penalty resulting in some coefficients being shrunk to zero and increasing parsimony.The table below describes the number of non-zero coefficients and their percentage of the 2500-dimensional feature space.

R reports that Ridge has 2,497 (likely due to internal rounding) non-zero coefficients while Elastic Net and Lasso have 1,048 and 1,025 respectively. Thus, from our 2500 predictors, Elastic Net says we only need 41.92% of our feature space (1048/2500) to determine if a review is positive or negative with a reasonable level of confidence. Ridge, using the entire feature space, gains an additional 0.00112 in AUC but requires an additional 1,452 features to achieve this marginal improvement.

```
##        Method N.Non.Zero Perc.Of.Feature.Space
## 1 Elastic Net       1048                0.4192
## 2       Lasso       1025                0.4100
## 3       Ridge       2497                0.9988
```







Part IV - Type I & Type II Error Rates

We now consider the Type 1 and Type 2 Error Rates for Elastic Net, Lasso, and Ridge in both Testing and Training Data. We immediately see that Ridge outperforms the others in Type 1 Error but it has the largest Type 2 Error rates. Given it uses 100% of the feature space, its lack of parsimony and high Type 2 Error rates make it a sub-optimal model.

While Elastic Net & Lasso perform equally well for Type 1 Error in training and testing, Lasso performs marginally better when it comes to Type 2 Error. Elastic Net is the preferred model here given its greater parsimony and better Test AUC.

```
##   Method Threshold Type.1.Train Type.1.Test Type.2.Train Type.2.Test
## 1  Elnet       0.5      0.02760     0.04232      0.22175     0.30350
## 2  Lasso       0.5      0.02840     0.04432      0.21725     0.29875
## 3  Ridge       0.5      0.01368     0.02368      0.31075     0.41425
```

Part V - Minimum Type I & Type II Error Rate Difference

We identified the thresholds for which the absolute difference between the Type 1 and Type 2 error rates was a global minimum. For the training data, the threshold of 0.3 was the same for Elastic Net, Lasso, and Ridge. In the test data, the threshold was 0.27 for all three.

```
##   Method Threshold.Train Min.Diff.Train Threshold.Test Min.Diff.Test
## 1  Elnet             0.3        0.00152           0.27       0.00083
## 2  Lasso             0.3        0.00259           0.27       0.00119
## 3  Ridge             0.3        0.00273           0.27       0.00287
```