Introduction

* Data set is binomial and small
* Use elastic net on a binomial model. Can do using package glmnet

Body

* Pick which functions I shall use

Conclusion

* Type of outputs I can expect and what interpretations I can make

References

Introduction

This introduces a bias term to the MSE so even if the variance of a dataset is large, the bias can be used to reduce the MSE leading to an increase in prediction accuracy (Tishbirani, 1996).

Ridge regression can only be used if the design matrix has full rank, n > p. If p > n than LASSO has to be used. Also, until ridge regression, LASSO does not just shrink coeffects it actually reduces them to zero. In practice this can been seen as starting with a dataset of 80 features and after applying the LASSO only 5 remain with coefficients great than zero. This makes it a very powerful tool for interpreting data sets with large feature count to determine which are the truly important features.

LASSO does have its own limitations such as if n > p but the features have high correlation than ridge regression will still outperform the LASSO (Zou and Hastie, 2005).

Elastic net is a regulation technique that has been presented to overcome the short comings of LASSO regression. Instead of a norm as a MSE penalty term it uses the elastic net penalty **Equation here.** It is an amalgamation of ridge and LASSO penalties controlled by the variable alpha. When alpha is 0 it becomes pure ridge regression. When alpha is 1 is pure LASSO. By choosing an alpha value in between it maximizes the benefits of both regression methods.

The data set I will be using for the final project is a modification of the wine quality data sets presented by [Cortez et al., 2009]. The red wine data and the white wine data where concatenated into a single dataset and an addition attribute colour was added to the new data set indicating which original data set the observations came from. This the attribute or feature will be used as the output for the analysis making it a binomial regression problem. When creating my training set will equal downsample both class for a total n of 100. Reducing the number of observation in a binomial data set increases the difficulty of the regression, making this an interesting test of the regression methods. (Hastie and Qian, 2016).

This dataset will be a good candidate for elastic net regularization because it will allow for feature reduction due to coefficients being shrunk to zero but without the LASSO limitations that occur when n > p and features are highly correlated as may be the case here. The R package Glmnet will be used to analysis this data set as it allows for use of the LASSO or the elastic net penalty to be used to fit a generalized linear model.

Body

Start off by splitting data into training and testing.

Functions I may use in the analysis:

* Will use cv.glmnet to perform a cross validation on my data to determine critical lambdas including lambda min the lambda that minimizes MSE. By default alpha=1 so LASSO is used here. Introduce elasticnet by adjusting alphas. Can use the argument foldid and run for a range of alpha to also use cross validation to select the best alpha. Because I have a binomial data set will measure misclassification error instead of MSE by specifying that in arguments.
* Once I have chosen a critical lambda and alpha I can use glmnet to fit a model these values, making sure to specify family binomial. This object can be used in the predict function across my testing data.

Conclusion

* As well as determining just trying to produce the best fitting model all the way I can plot various outputs in order to learn more about my dataset. By plotting the glmnet object for glmnet over all the lambdas I can see which of my coefficients shrunk to zero first giving me an idea of order of importance. My critical alpha value will give insight into how correlated the features are. We know when features are highly correlated ridge regression outperforms lasso and as alpha approach zero the elastic net is favoring ridge regression so we can infer the smaller the alpha the greater the correlation among features.