**LASSO Regression**

Regression models are commonly used in Machine Learning and Statistical Analyses. A standard regression model works by fitting a line through the data points that minimizes SSR (Sum of Squared Residuals). Unfortunately, applying standard regression methods to a generate a model tend to lead to overfitting to the training data and overestimation of how well the model performs. The model tends to perform poorly when making predictions on new data that is more extreme. Therefore, various penalization or regulation regression techniques can be used to solve this problem.

LASSO (Least Absolute Shrinkage and Selection Operator) regression is a shrinkage and feature selection method for regression models that address the problem. It identifies a subset of variables that minimizes the prediction error. This is achieved by constraining the model parameters, which shrinks some of the regression coefficients towards zero by forcing the absolute value of the coefficients to be less than λ, a fixed value that is optimized by a k-fold cross-validation approach. The variables with coefficient of zero will be excluded from the regression model. LASSO Regression has been shown to outperforms standards regression and solve the problems of overfitting and optimism bias.

**LASSO Regression to Predict the Mass of the Blackhole**

A data file has been generated from the Cloudy Model c17 (Ferland et al. 2017) that has 3 input parameters: Mass of the blackhole, Ionization Parameter (U), and Electron Density (Hden), and 55 outputs of the different diagnostic line ratios. The goal is to predict the mass of the blackhole from the dataset that is generate by the Cloudy Model. First, a LASSO regression model was used as a feature selection method to choose a set of features that are most sensitive to U and the mass of the blackhole. The following features 5 features were selected:

Mg7(9)/Mg4

Na6(8)/Na3

Fe13/Fe6(1.01)

Na6(14)/Na4(6)

Na6(14)/Na4(9)

Then, a LASSO regression model (λ=0.00001) was created to predict the value of U from the dataset as the following:

1.18 [Mg7(9)/Mg4] + 0.54 [Na6(8)/Na3] + 2.29 [Fe13/Fe6(1.01)] - 2.20 [Na6(14)/Na4(6)] + 0.81 [Na6(14)/Na4(9)] - 2.68

The model to predict U has mean squared error of 0.091. Once the value of U is predicted by the LASSO Regression, another LASSO regression model (λ=0.00001) is used to predict the mass of the blackhole depending on the U value. If the value of U is less than or equal to -1 and greater than -2, the following LASSO regression model is used:

-3.54[Mg7(9)/Mg4] + 7.43[Na6(8)/Na3] - 19.98[Fe13/Fe6(1.01)] + 8.84[Na6(14)/Na4(6)] - 2.62[Na6(14)/Na4(9)] + 18.08

This model has a mean squared error of 0.576. If the value of U is less than or equal to -2 and greater than -3, the following LASSO regression model is used:

2.42[Mg7(9)/Mg4] - 7.20 [Na6(8)/Na3] + 16.44 [Fe13/Fe6(1.01)] -47.29[Na6(14)/Na4(6)] + 25.30[Na6(14)/Na4(9)] + 14.31

This model has a mean squared error of 0.589. If the value of U is less than or equal to -3 and greater than -4, the following LASSO regression model is used:

-27.76[Mg7(9)/Mg4] - 24.84[Na6(8)/Na3] + 50.51[Fe13/Fe6(1.01)] - 104.66[Na6(14)/Na4(6)] -6.97[Na6(14)/Na4(9)] - 47.33

This model has a mean squared error of 0.441.

**Resources:**

Ranstam J , Cook JA . Statistical models: an overview. *Br J Surg* 2016; 103: 1047.