EE-622 Assignment -1

Report

Name: K.Samanth Reddy

Roll no:130102031

I have taken first 135 patients for training purpose and tested on the remaining set of the data.I have assumed the NA values in the data to zero. I have used the glmnet library and found out the accuracies for models .

Model	Alpha	Accuracy
RIDGE	0	59
ELASTIC NET	0.5	68.22
LASSO	1	60.41
RANDOMFOREST		73.4375

Cross validation involves portioning the data into complimentary subsets and performing the analysis on one subset and performing the analysis on the testing set.

For Ridge model:

Ridge Regression is a technique for analyzing multiple regression data that suffer from multicollinearity

n-folds	Accuracy
5	46.875
8	63.54
10	68.75
15	61.45

For elastic net:

Elastic net is a regularized regression method that linearly combines the L1 and L2 penalties of the lasso and ridge methods.

n-folds	Accuracy
5	50.52
8	62.58
10	60.93
15	65.625

For LASSO:

 LASSO is a regression analysis method that performs variable selection and regularization in order to enhance accuracy.

n-folds	Accuracy
3	46.875
5	58.33
10	60.41
12	65.41

For cross validation I have found out the minimum lambda for 1 in each model and found out the accuracies.

Model	Minimum lambda	Accuracy
Ridge	8.27	62.5
Elastic net	0.11147	60.93
Lasso	0.073	60.41

The coefficients for the regression model are found out using the coef function and genes for which the coefficient are negligible are considered to be non important genes ,in the sense that they won't contribute significantly to the regression model and the high coefficient values affect significantly to the model.Ridge regression addresses some of the problems of ordinary least

squares by imposing a penalty on the size of coefficients. Lasso is a linear model that estimates sparse coefficients. It is useful in some contexts due to its tendency to prefer solutions with fewer parameter values, effectively reducing the number of variables upon which the given solution is dependent.

Airway PI3K Pathway Activation Is an Early and Reversible Event in Lung Cancer Development**:**

One approach to assess signaling pathway activity uses the gene expression pattern resulting from the in vitro activation of the pathway to predict its activity in patient samples. To predict pathway activation in normal airway microarray samples, genomic signatures are used to train a binary regression model for each of the studied pathways and then applied these signatures to an external gene expression data set to determine respective pathway activity. whether the increased PI3K pathway activation in normal airway cells is specifically associated with lung cancer or is the result of other potentially confounding factors, such as differences in cumulative smoke exposure.