Comparison between different linear regression models

Models	# of predictors in the model	Mean absolute error
OLS	56	3147.012
Forward selection	56	3147.012
Stepwise selection	8	3149.006
Ridge regression	56	3145.607
LASSO	35	3133.187
SCAD	21	3136.644

Comparison between different classification methods

Classifier	Accuracy	Sensitivity	Specificity	AUC
Penalized (LASSO) logistic regression	0.6271	0.29	0.90	0.6990
Support vector machines (SVM)	0.6457	0.56	0.72	0.6858
Random forest	0.6584	0.66	0.66	0.7177

Distribution of Shares

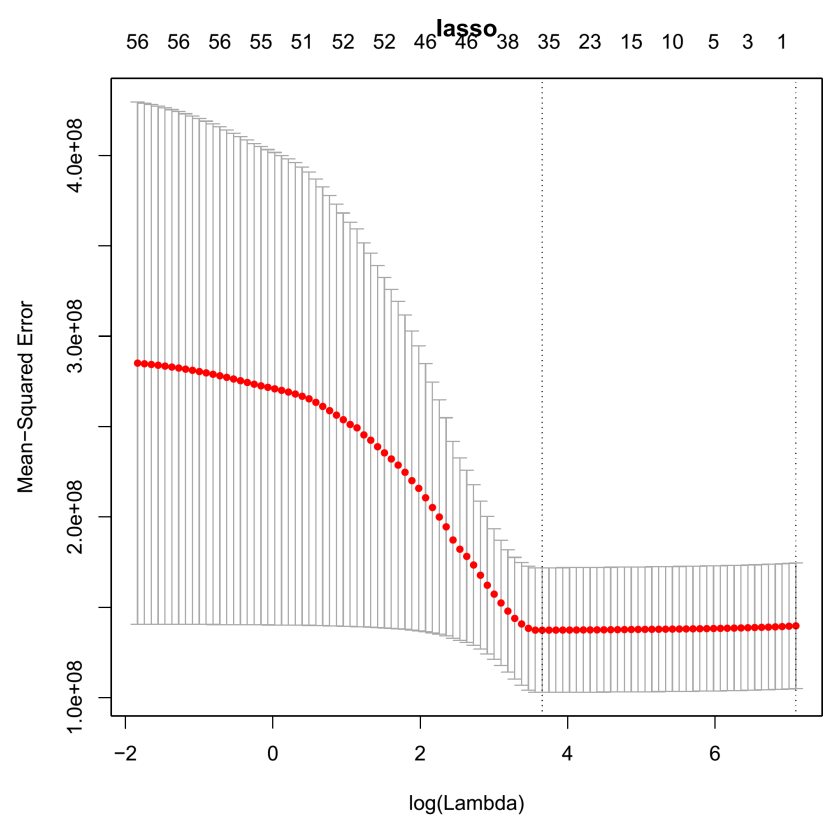| Number of Shares	Number of Observations
Less than 1,500	20,082
Between 1,500 & 10,000	17,347
Between 10,000 & 50,000	2,011
Between 50,000 & 100,000	146
Over 100,000	58

Distribution of Shares
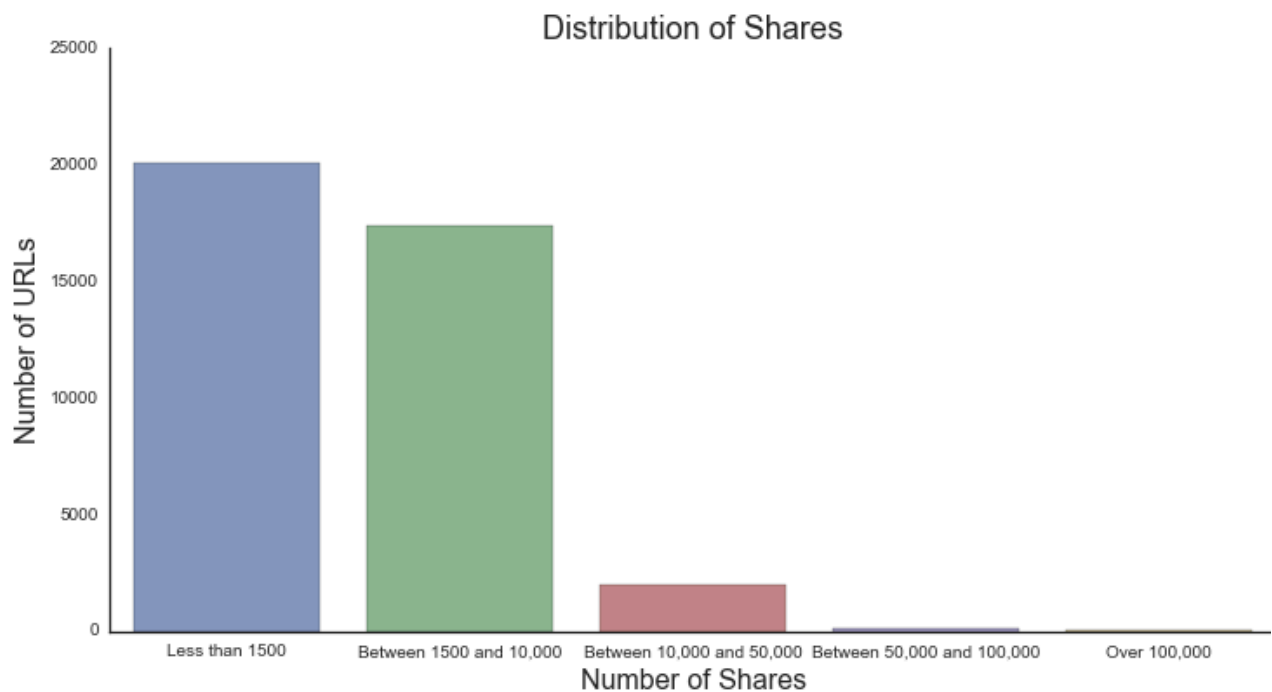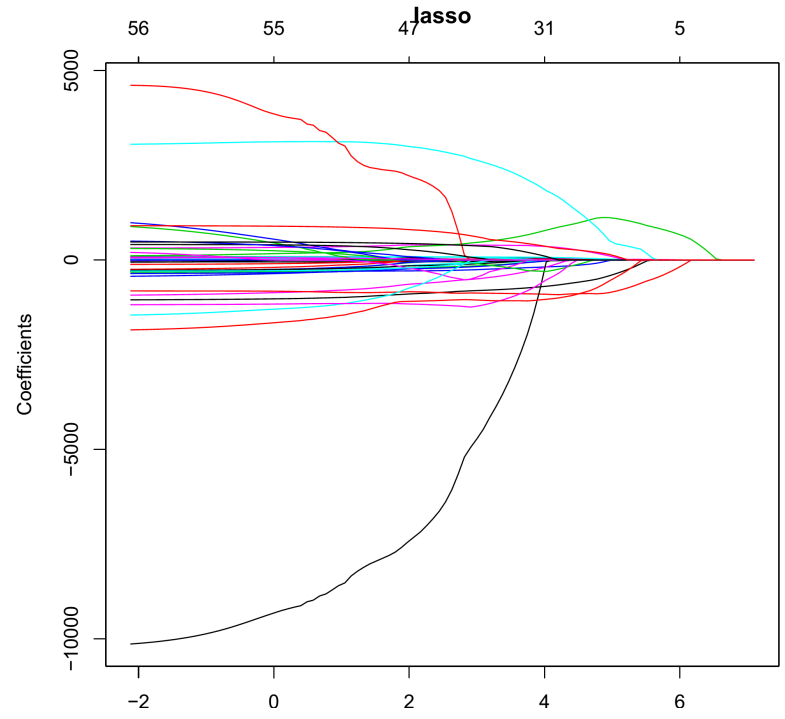
Ten-fold cross-validation MSE for ridge
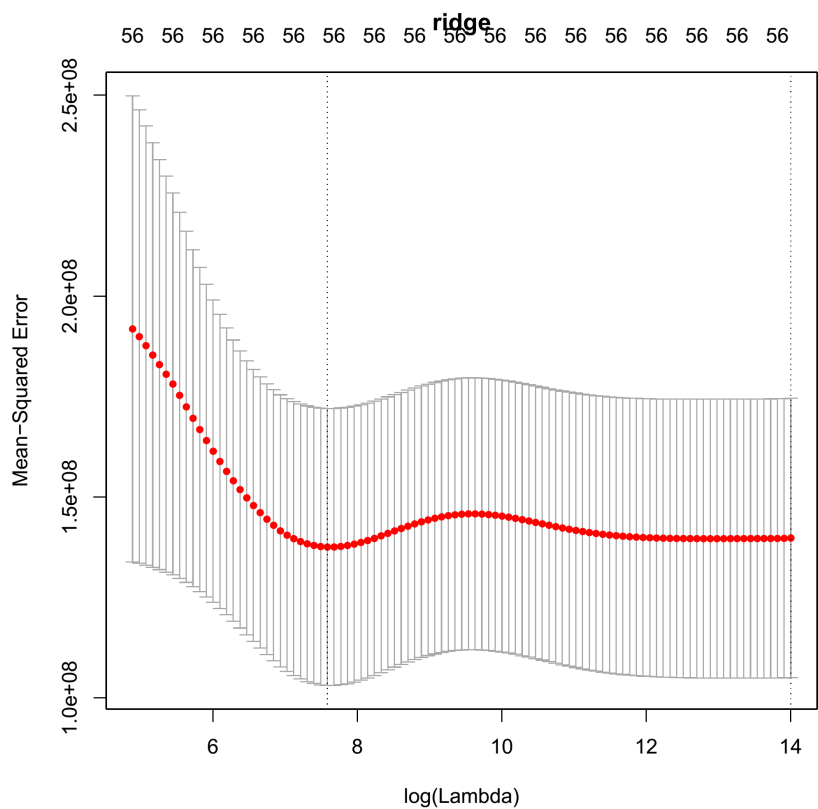
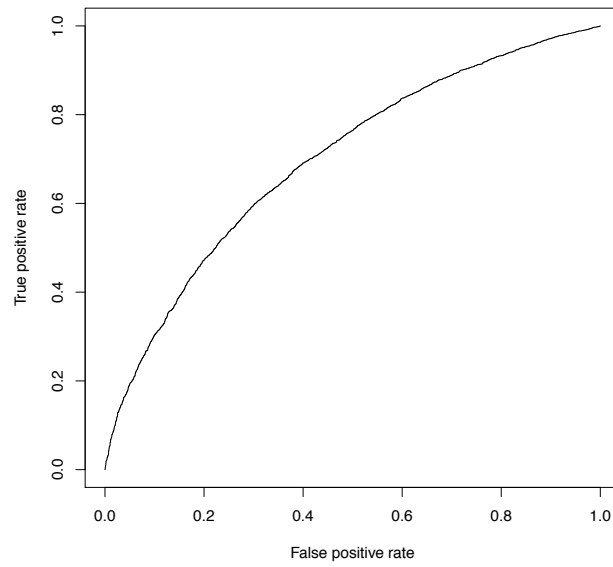# Display of corresponding ridge coefficient estimates
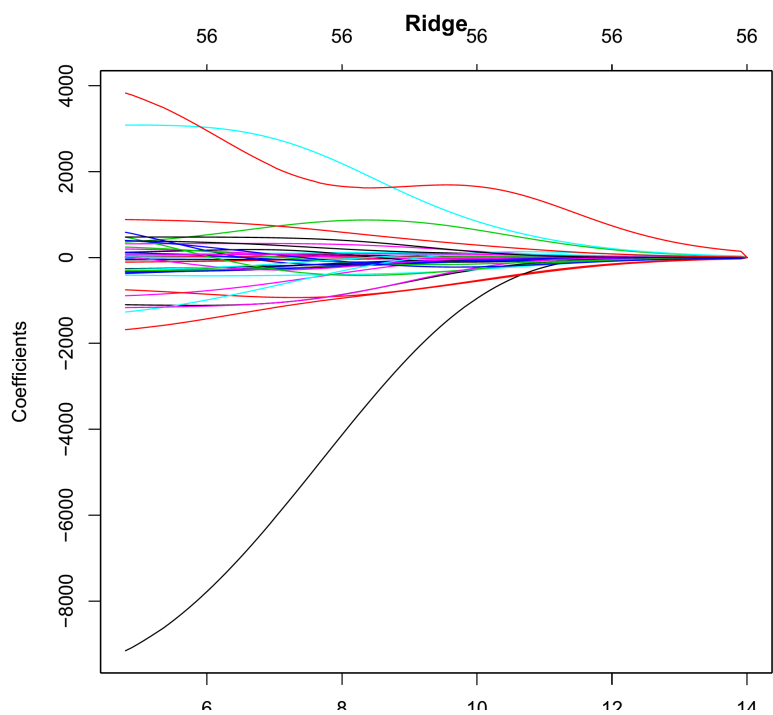
Ten-fold cross-validation MSE for lasso
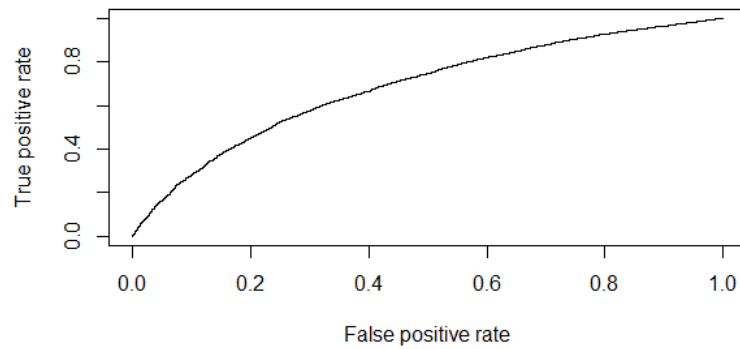

Display of corresponding lasso coefficient estimates

Classification- ROC Curve for Logistic Regression

Classification- ROC Curve for SVM

Classification- ROC Curve for Random Forest