# Evaluate testing data (binary-class) - Lasso

*John Doe*

*2019-01-30*

## Contents

Note: The two differences between Lasso and Tree-based methods are:

1. Lasso has its own inherent feature selection process.
2. Lasso's vimp will be based on how many times the feature exist in all runs. Regression coefficients may be presented for binary outcomes
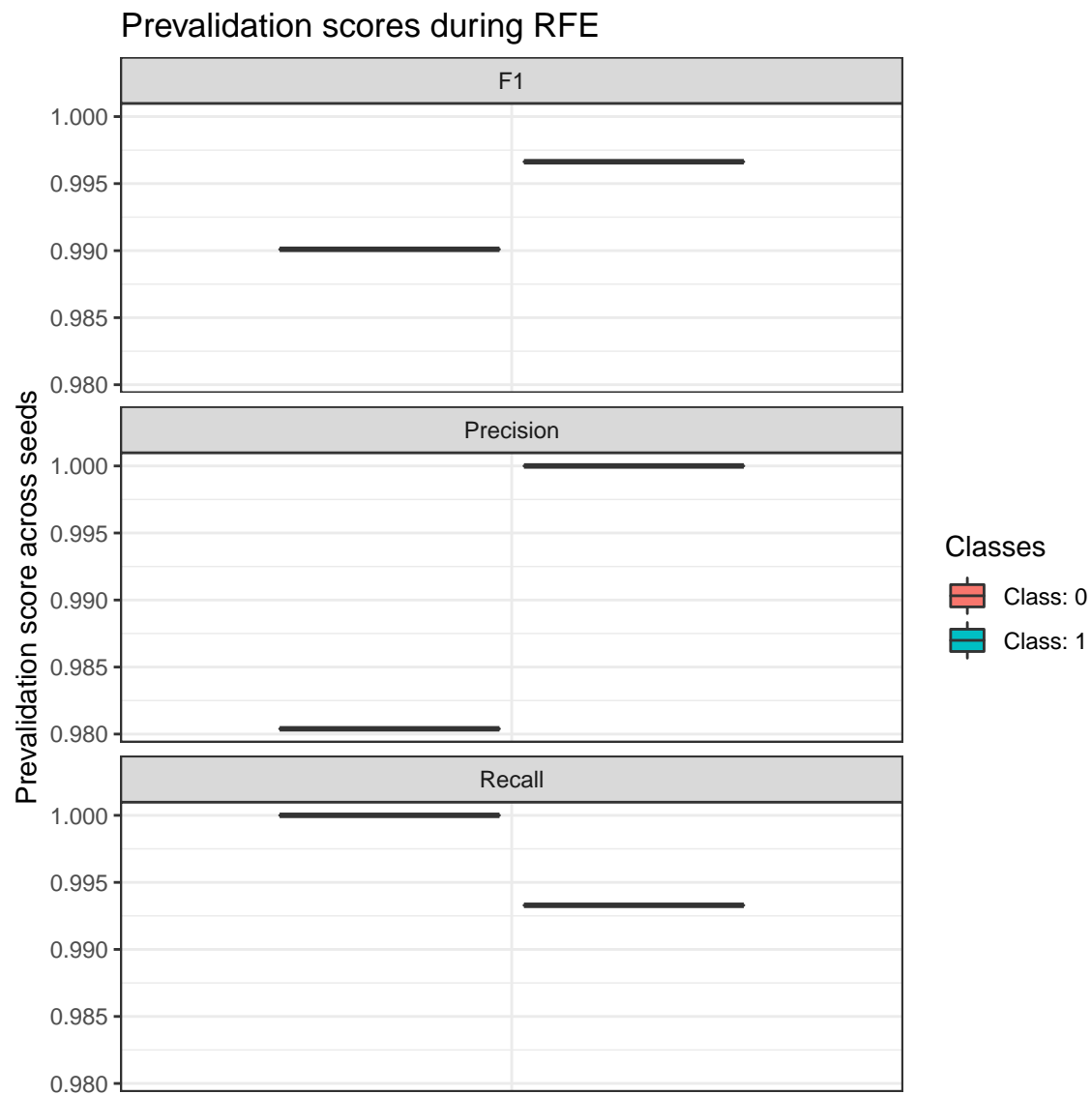
```
## user input
project_home <- "~/EVE/examples"
project_name <- "lasso_binary_outCV_test"
```

## 0. Load Data

```
## 199 of samples were used
```

```
## 100 of full features
```

```
## 4 runs, each run contains 3 CVs.
```

```
## Labels:
```

```
##
##    0   1
##   50 149
```

# 1. Scores

## 1.1 Scores per Class

### Prevalidation scores during RFE
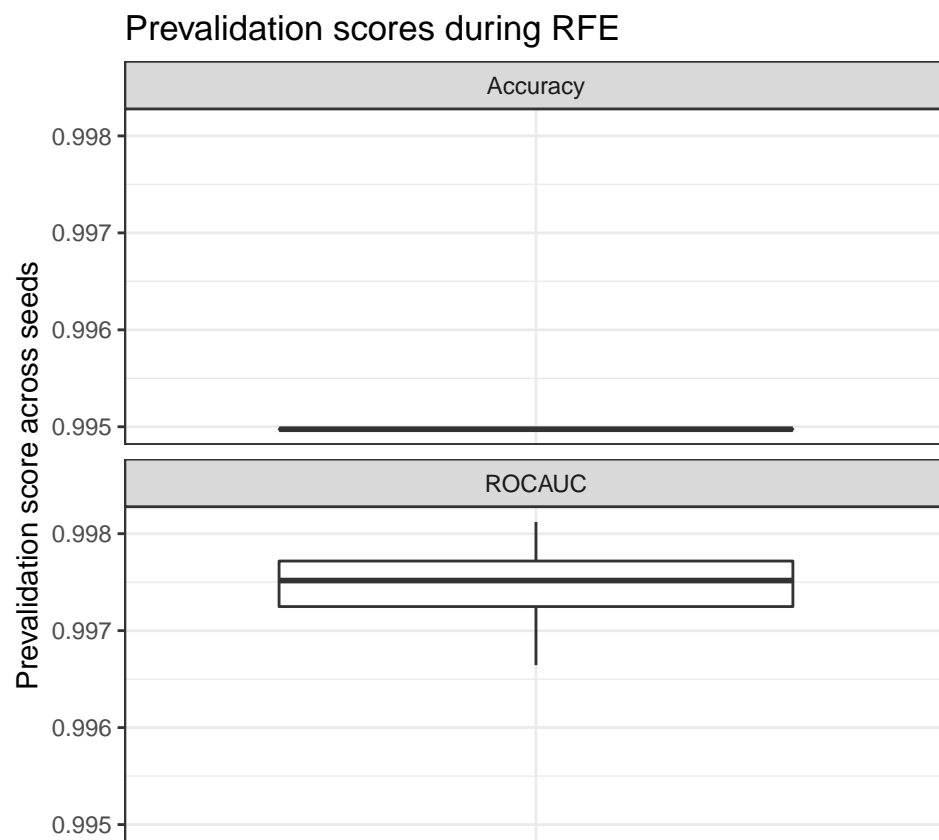


Confusion Matrix

```
## confusion matrix at feature size = 100
## sum across 4 seeds

##           Reference
## Prediction   0   1
##          0 200   4
##          1   0 592
```
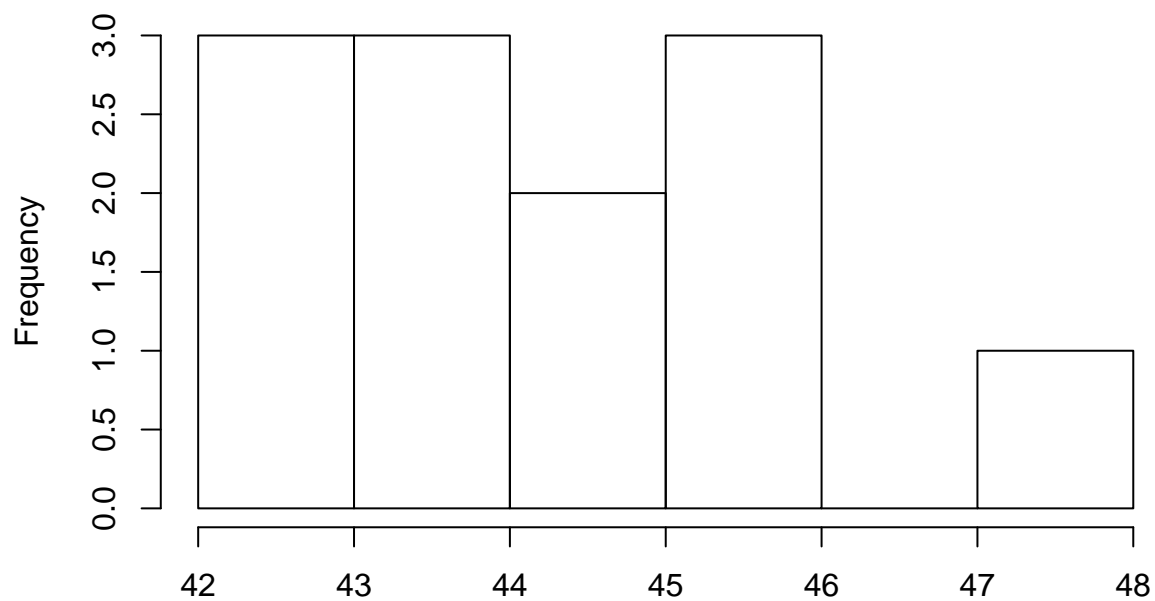
**1.2 Average score**

## Prevalidation scores during RFE

## 2. Important Features
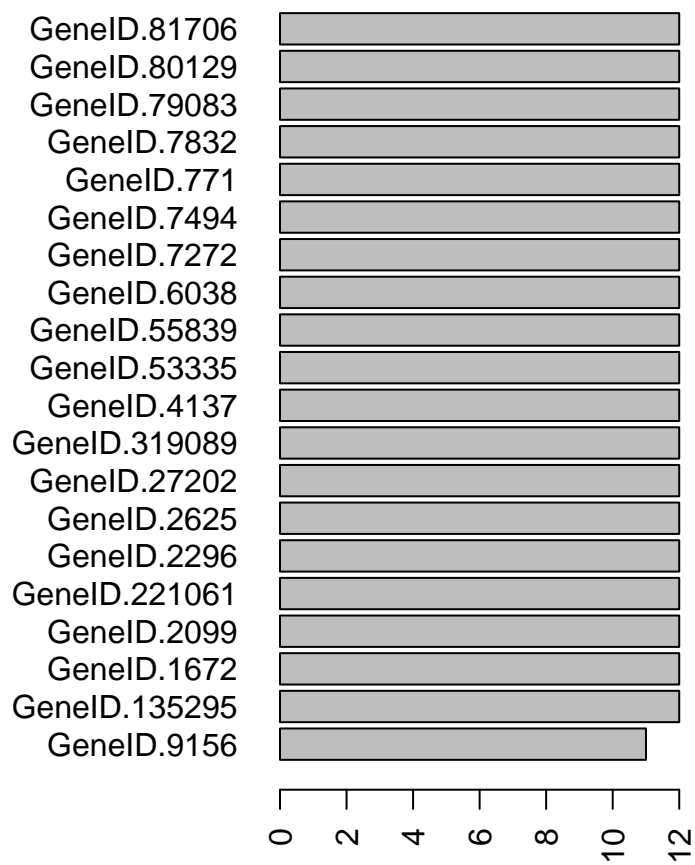
**distribution across 4 seed x 3 CV**
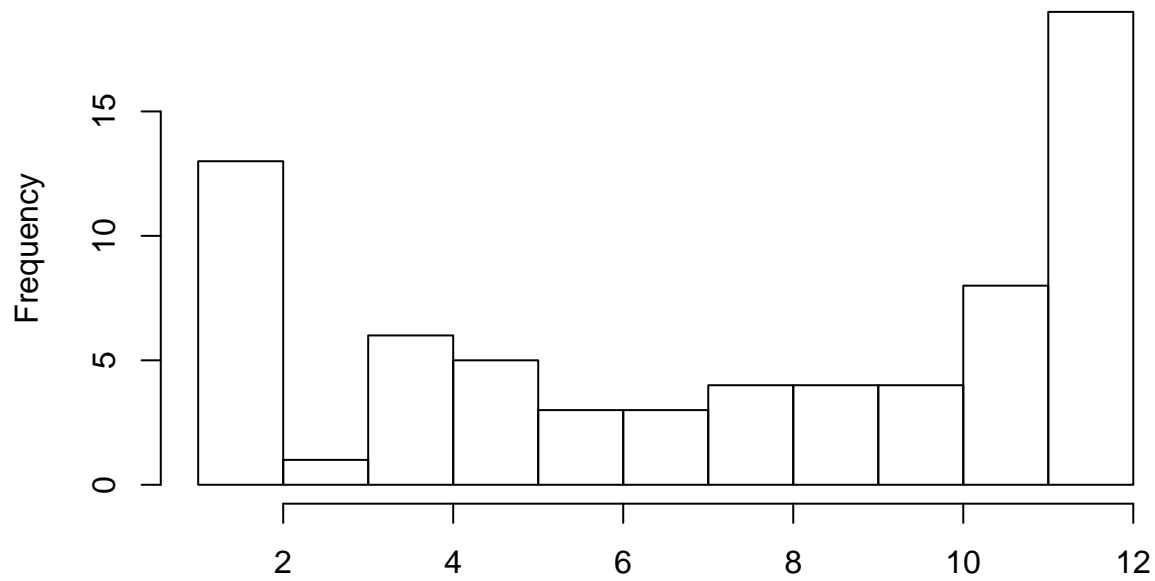


# of times a feature is selected by lasso (alpha= 0.5 )

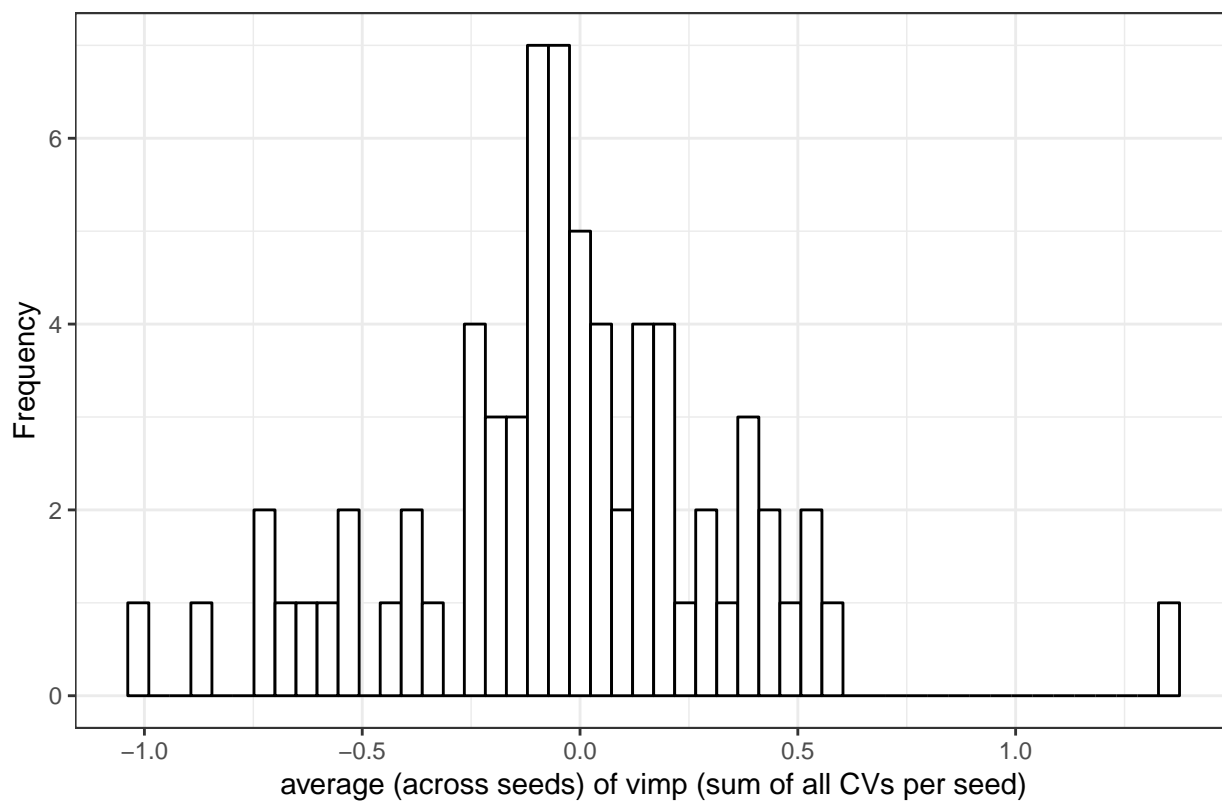**summary of model size across 12 runs**



number of retained features in one run

## Number of times a feature is use



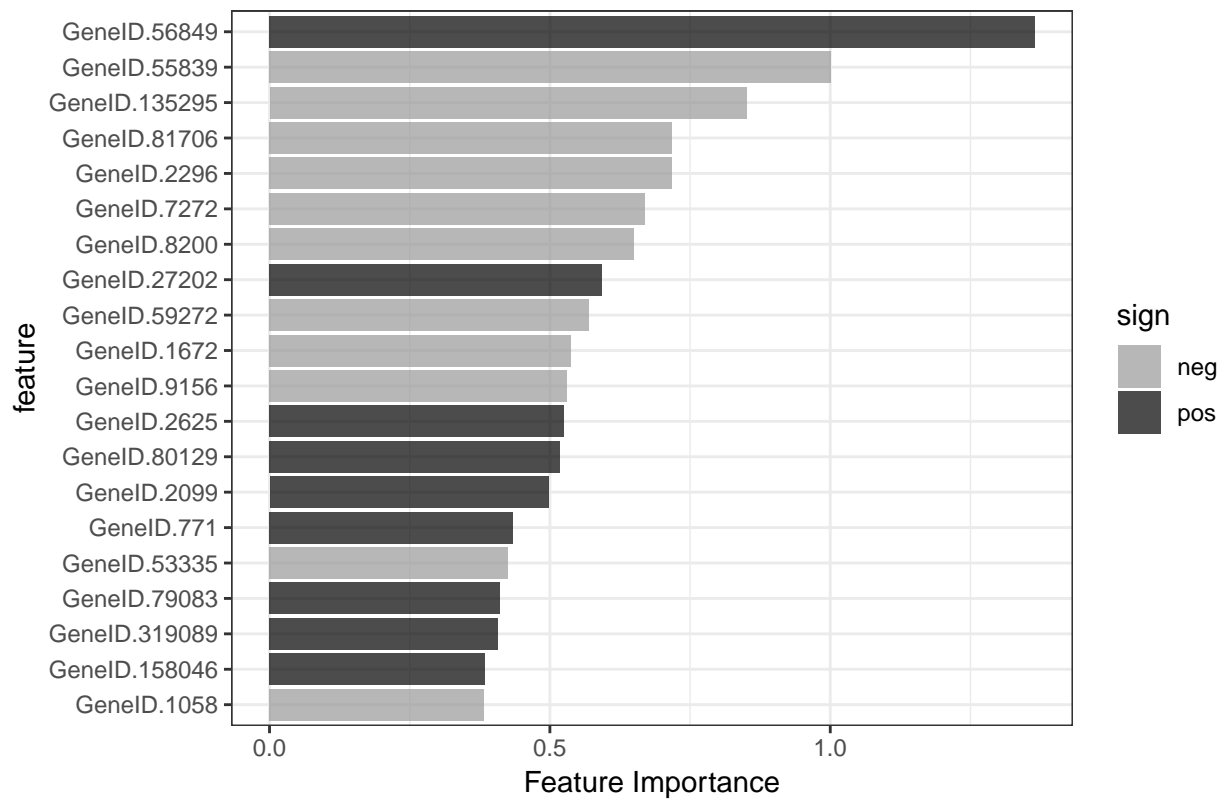(currently only Lasso has this graph)

## Distribution of Feature Coefficient



## Top Features

**Heatmap of top 20 important features**