Evaluate testing data (binary-class) - Lasso

EVE W.

2019-11-16

Contents

0. Load Data 1
1. Scores 2
    1.1 Scores per Class 2
    1.2 Average score 3
2. Important Features 4

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.
```

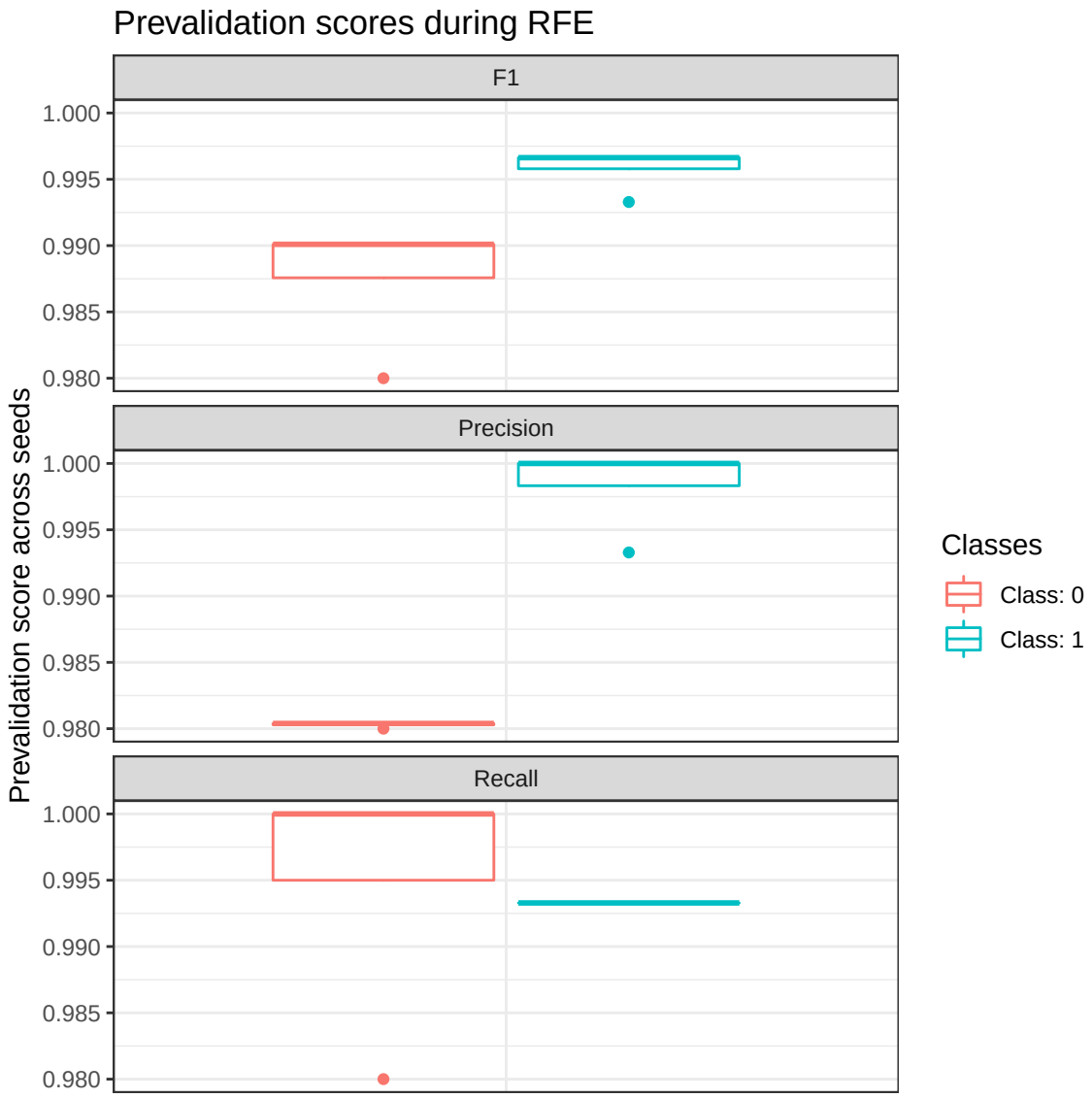
```
## Labels:
```

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

run with lasso.r.

1. Scores

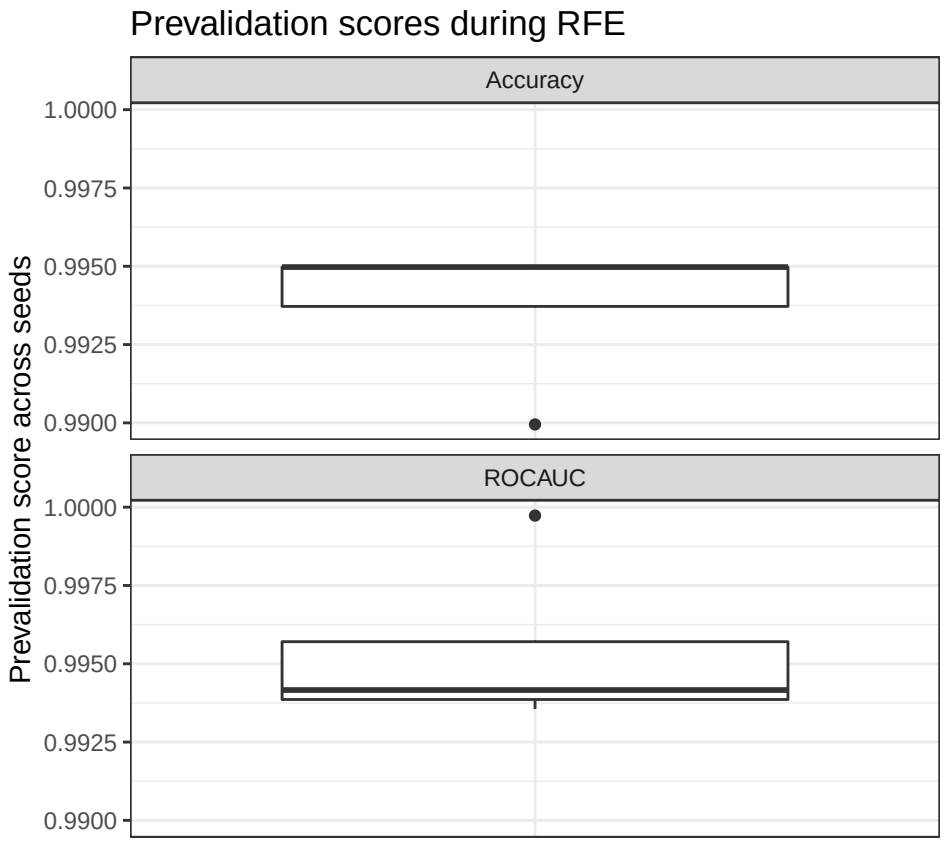
1.1 Scores per Class



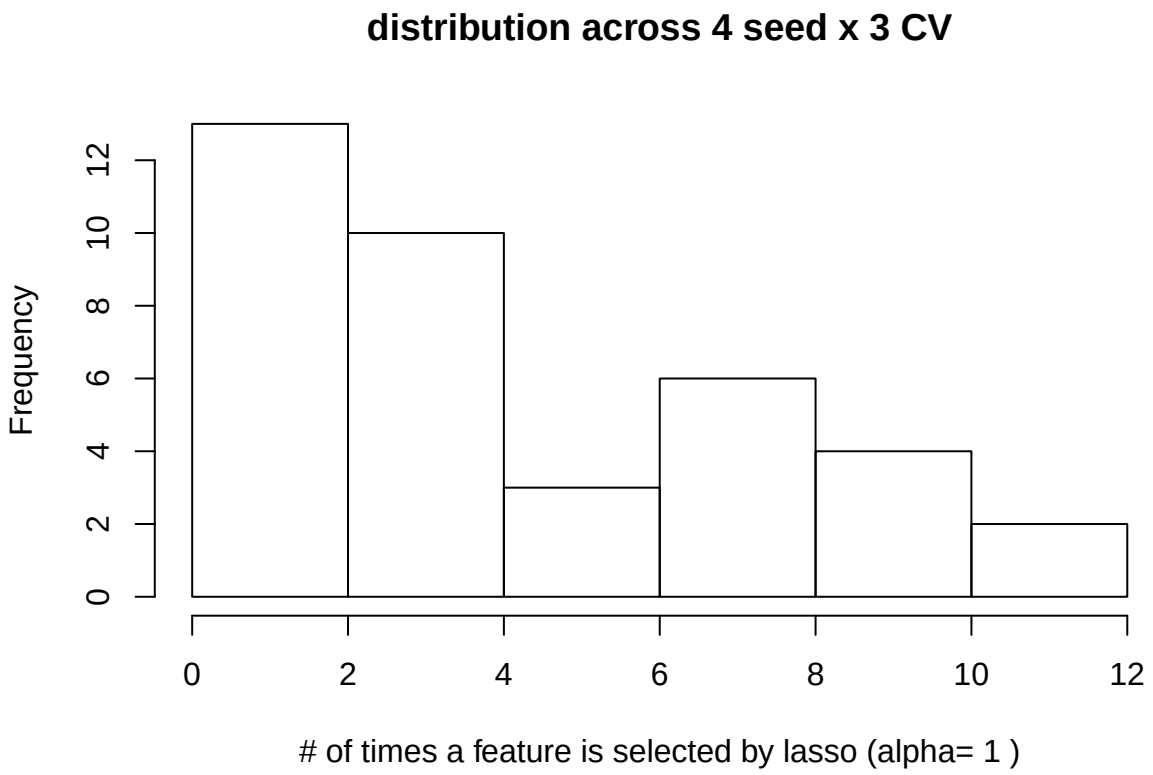
Confusion Matrix

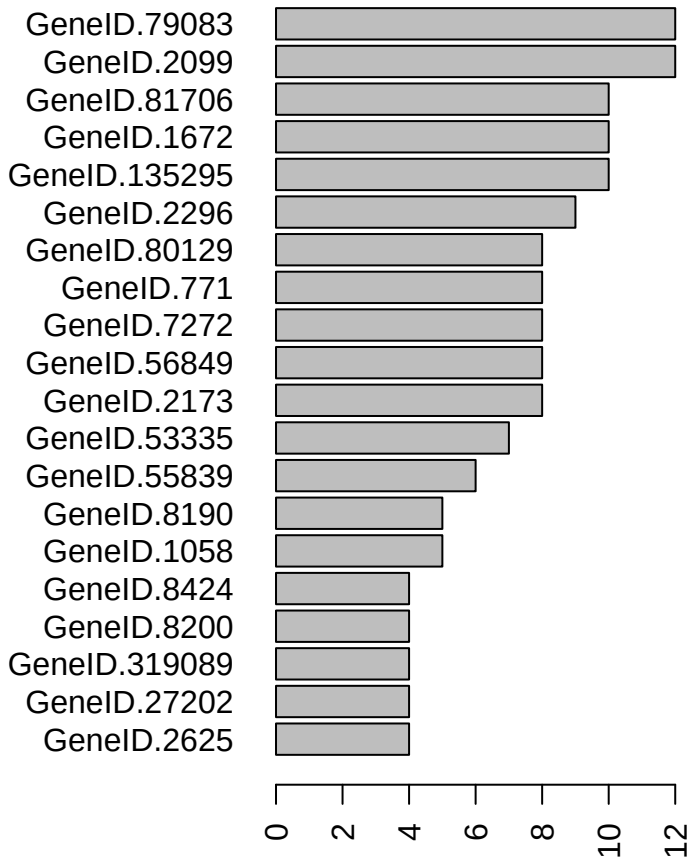
```
## confusion matrix at feature size = 100
## sum across 4 seeds

##            Reference
## Prediction   0   1
##          0 199   4
##          1   1 592
```

1.2 Average score

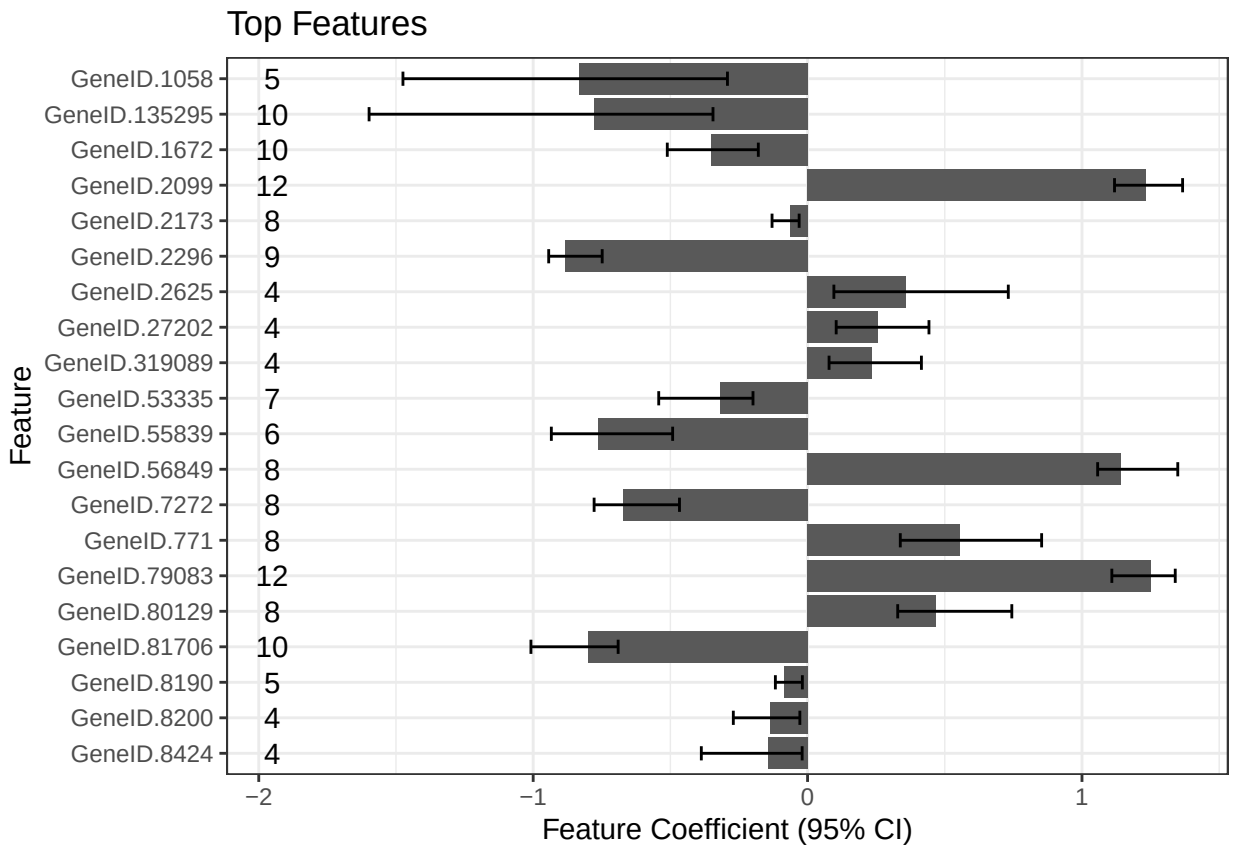


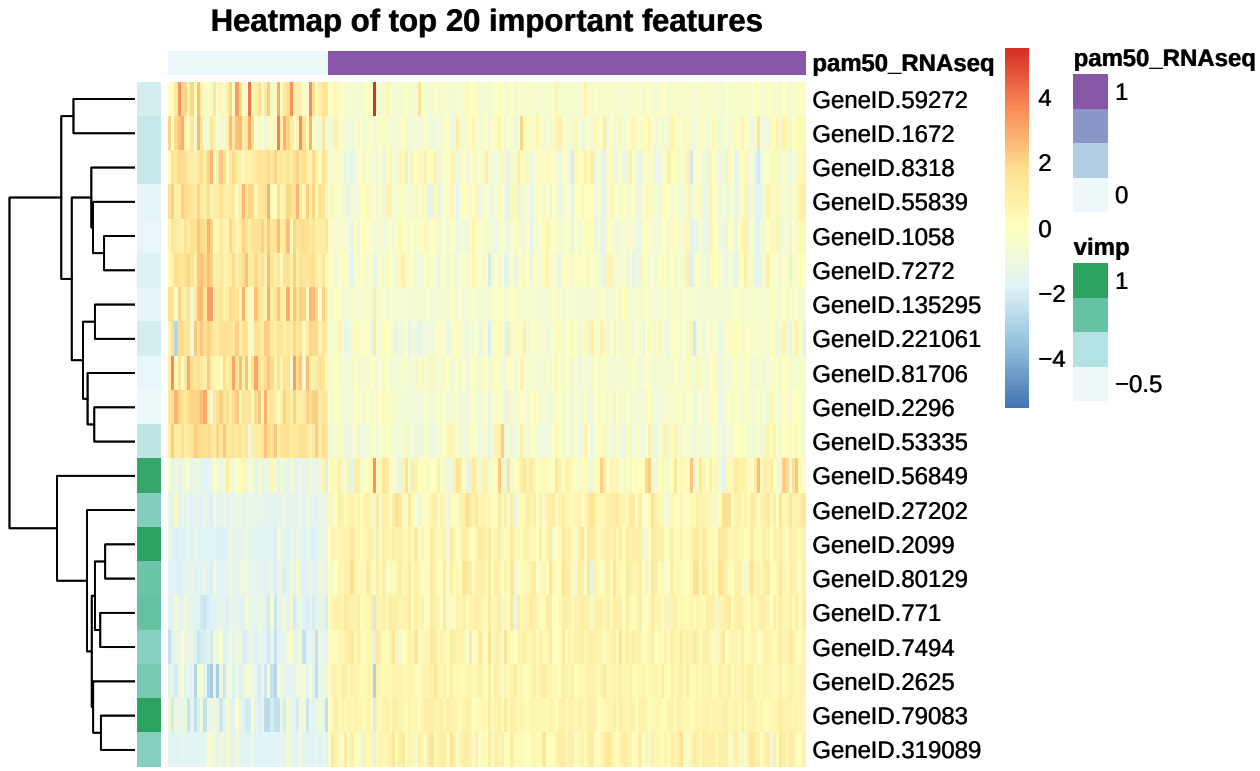
2. Important Features






```
## (currently only Lasso has this graph)[1] "there are 38 unique features used from the 100 feature set
## [1] "summary of numer of features used in 4 seeds and 3 CVs"

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   11.00   13.75   14.50   14.83   16.00   19.00
```


Top Features
GeneID.1058
GeneID.135295
GeneID.1672
GeneID.2099
GeneID.2173
GeneID.2296
GeneID.2625
GeneID.27202
GeneID.319089
GeneID.53335
GeneID.55839
GeneID.56849
GeneID.7272
GeneID.771
GeneID.79083
GeneID.80129
GeneID.81706
GeneID.8190
GeneID.8200
GeneID.8424
5
10
10
12
8
9
4
4
4
7
6
8
8
8
12
8
10
5
4
4
Feature
−2
−1
0
1
Feature Coefficient (95% CI)


Heatmap of top 20 important features
pam50_RNAseq
GeneID.59272
GeneID.1672
GeneID.8318
GeneID.55839
GeneID.1058
GeneID.7272
GeneID.135295
GeneID.221061
GeneID.81706
GeneID.2296
GeneID.53335
GeneID.56849
GeneID.27202
GeneID.2099
GeneID.80129
GeneID.771
GeneID.7494
GeneID.2625
GeneID.79083
GeneID.319089
4
2
0
−2
−4
pam50_RNAseq
1
0
vimp
1
−0.5