Evaluate testing data (binary-class) - Lasso

EVE W.

2019-07-30

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_binary2"
```

0. Load Data

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

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

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

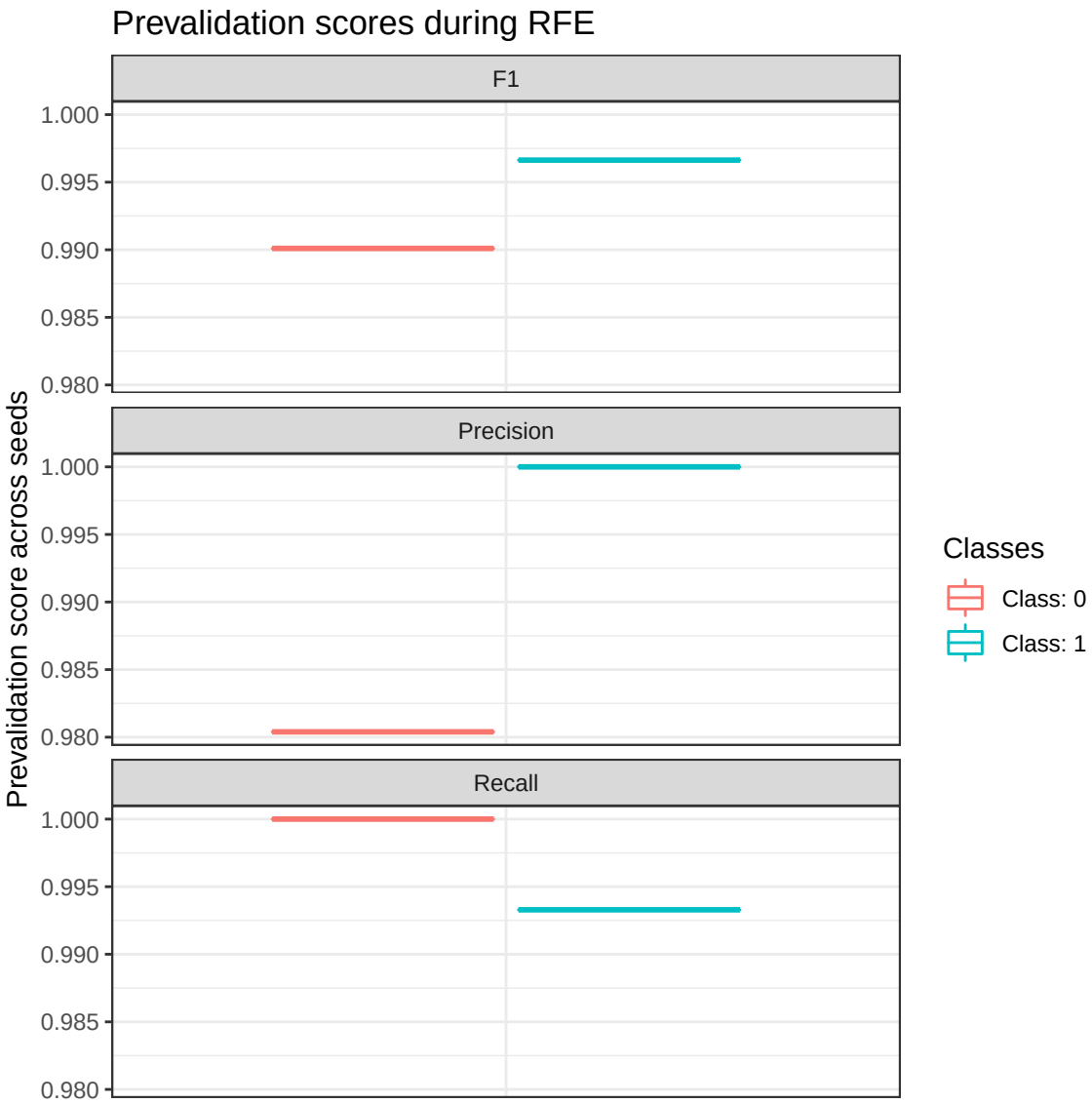
```
## Labels:
```

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

run with lasso.r.

1. Scores

1.1 Scores per Class

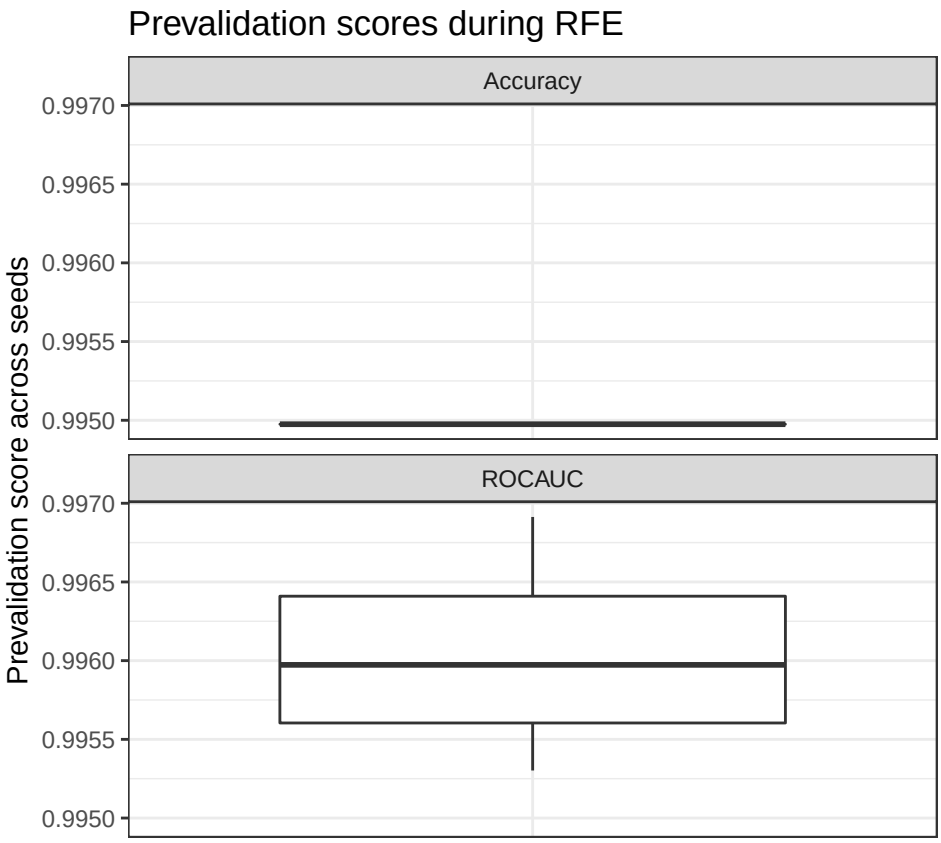


Confusion Matrix

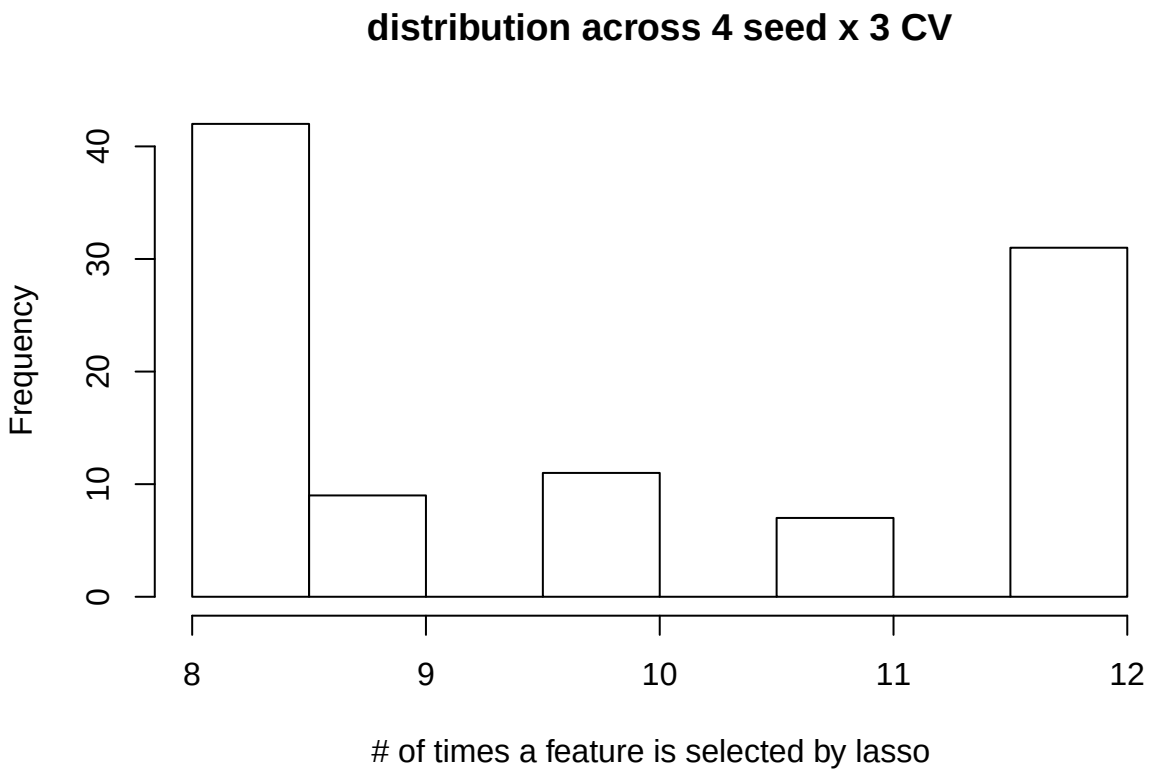
```
## confusion matrix at feature size = 100
## sum across 4 seeds
```

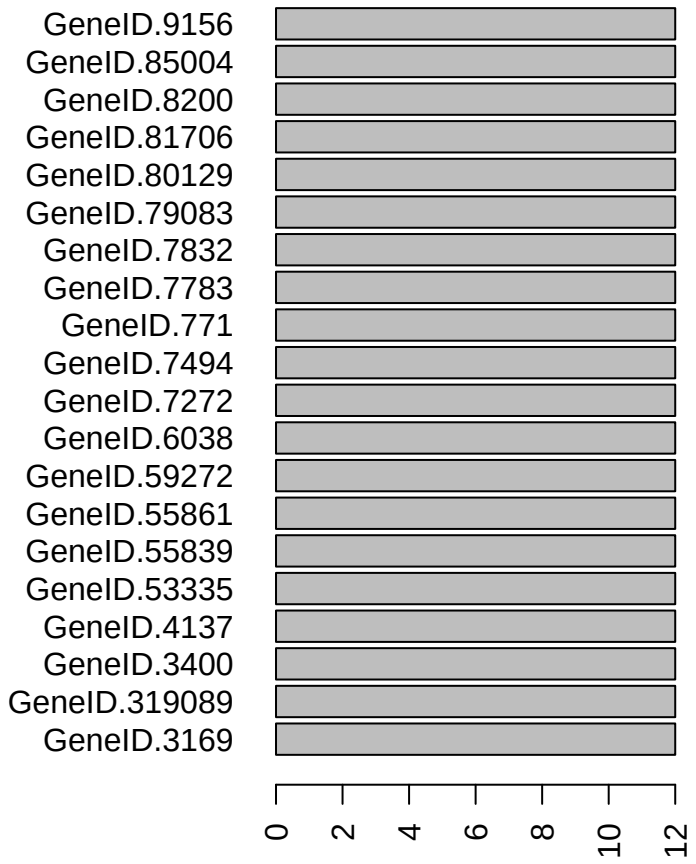
```
##            Reference
## Prediction   0   1
##          0 200   4
##          1   0 592
```

1.2 Average score



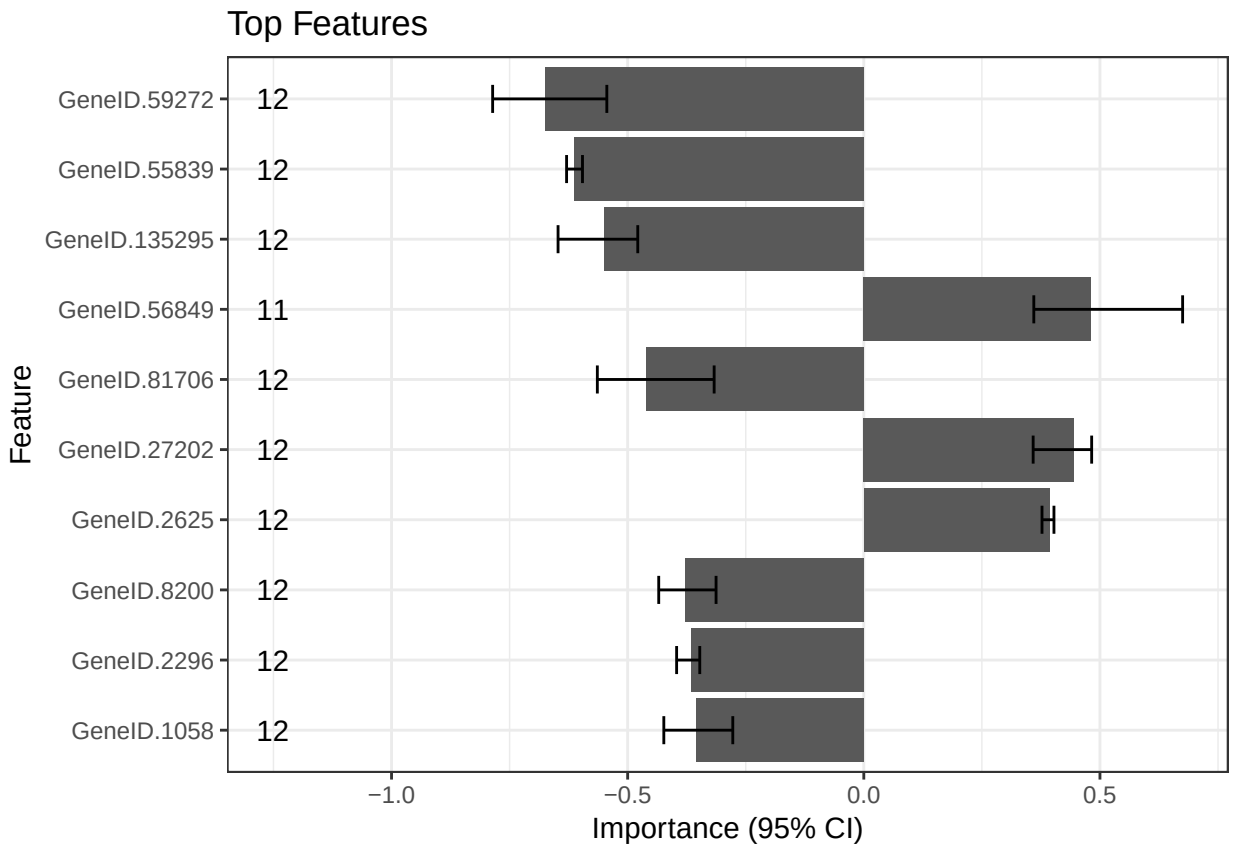
2. Important Features






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

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   42.00   47.00  100.00   81.33  100.00  100.00
```


Top Features
GeneID.59272
GeneID.55839
GeneID.135295
GeneID.56849
GeneID.81706
GeneID.27202
GeneID.2625
GeneID.8200
GeneID.2296
GeneID.1058
12
12
12
11
12
12
12
12
12
12
Feature
−1.0
−0.5
0.0
0.5
Importance (95% CI)

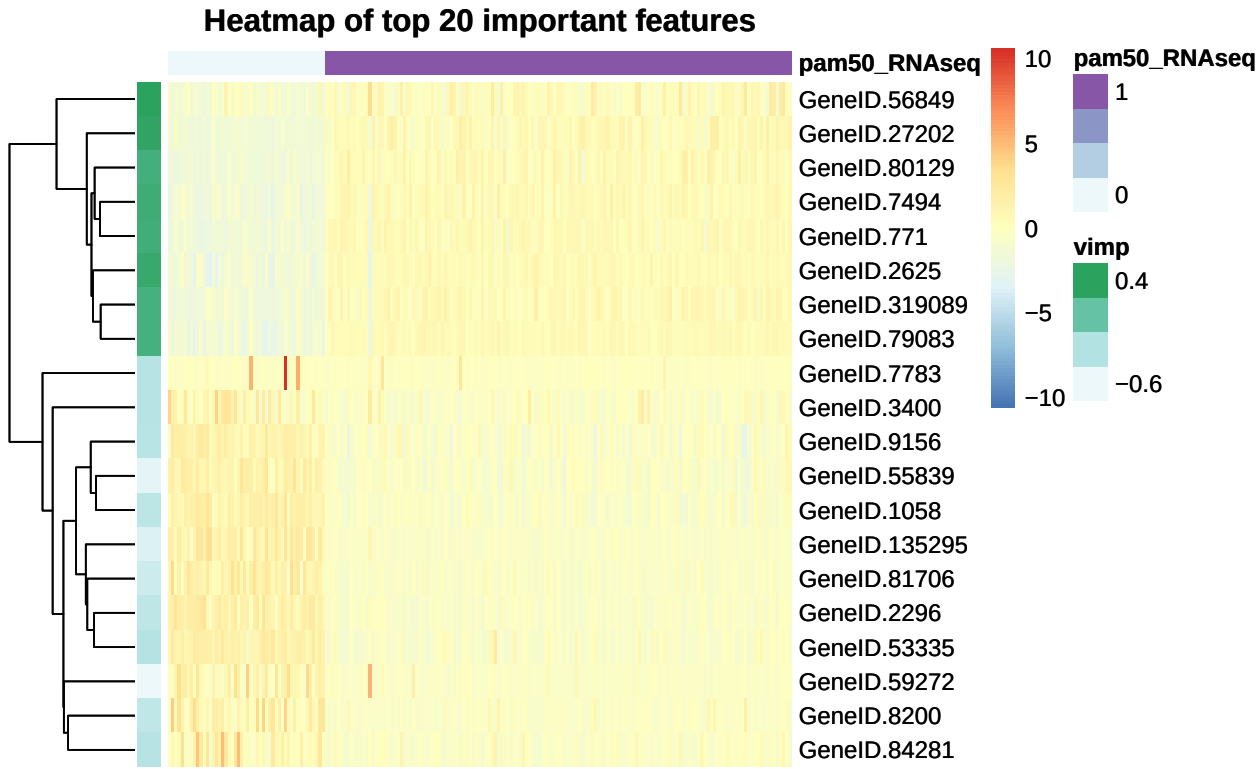

Heatmap of top 20 important features
pam50_RNAseq
GeneID.56849
GeneID.27202
GeneID.80129
GeneID.7494
GeneID.771
GeneID.2625
GeneID.319089
GeneID.79083
GeneID.7783
GeneID.3400
GeneID.9156
GeneID.55839
GeneID.1058
GeneID.135295
GeneID.81706
GeneID.2296
GeneID.53335
GeneID.59272
GeneID.8200
GeneID.84281
10
5
0
−5
−10
pam50_RNAseq
1
0
vimp
0.4
−0.6

Table 1: parameter selection

seed	alpha	lambda	cv
1001	0.5	0.0006179	1
1001	0.0	0.0436259	2
1001	0.0	0.0419874	3
1002	0.0	0.0427831	1
1002	0.5	0.0006208	2
1002	0.0	0.0422317	3
1003	0.0	0.0427720	1
1003	0.5	0.0005691	2
1003	0.0	0.0426477	3
1004	0.0	0.0429376	1
1004	0.0	0.0426406	2
1004	0.5	0.0005769	3