Evaluate testing data (binary-class) - Lasso

EVE W.

2019-11-16

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 5 CVs.
```

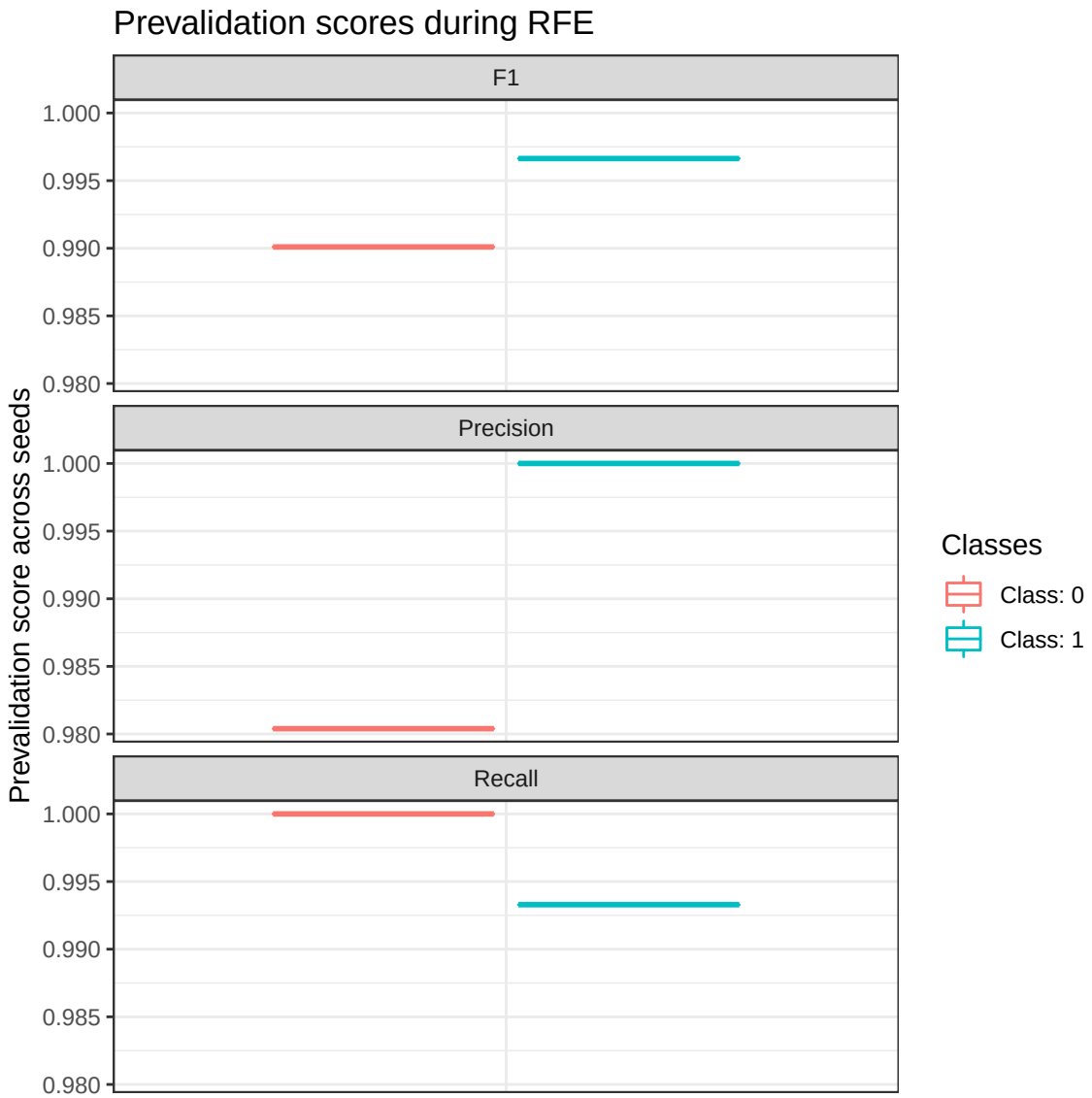
```
## Labels:
```

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

run with lasso.r.

1. Scores

1.1 Scores per Class

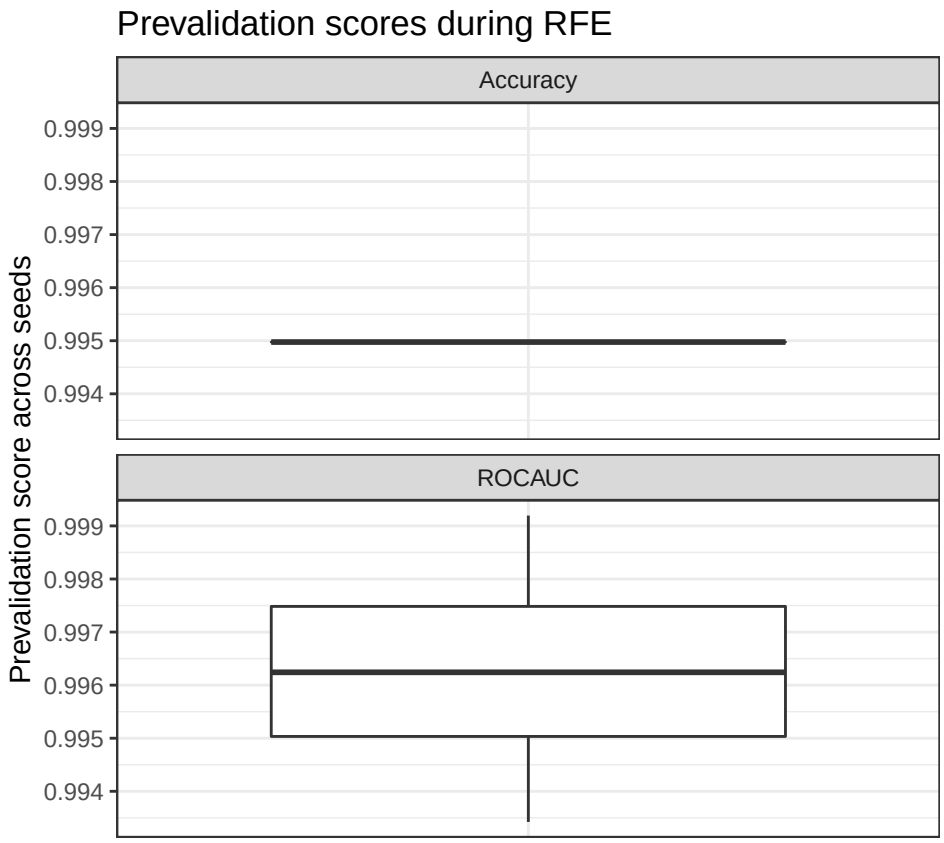


Confusion Matrix

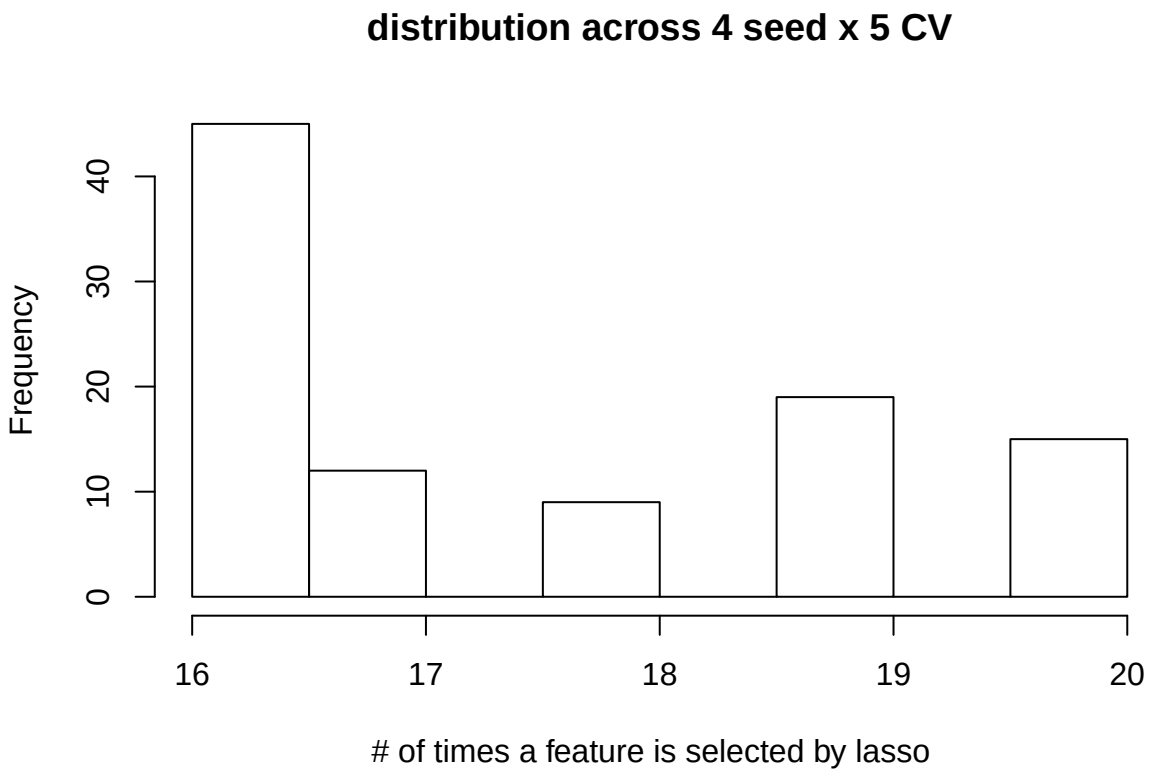
```
## confusion matrix at feature size = 100
## sum across 4 seeds
```

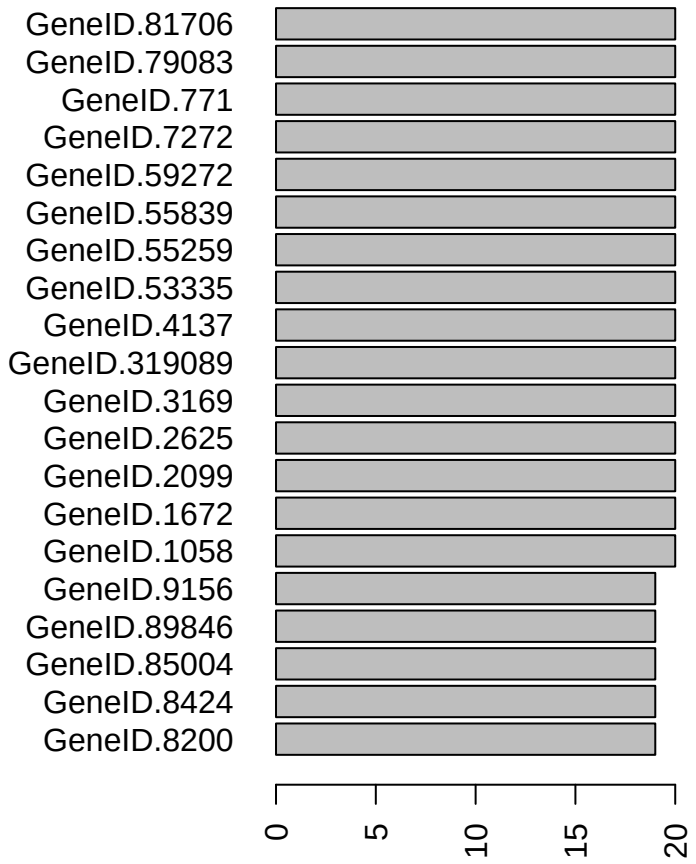
```
##            Reference
## Prediction   0   1
##          0 200   4
##          1   0 592
```

1.2 Average score



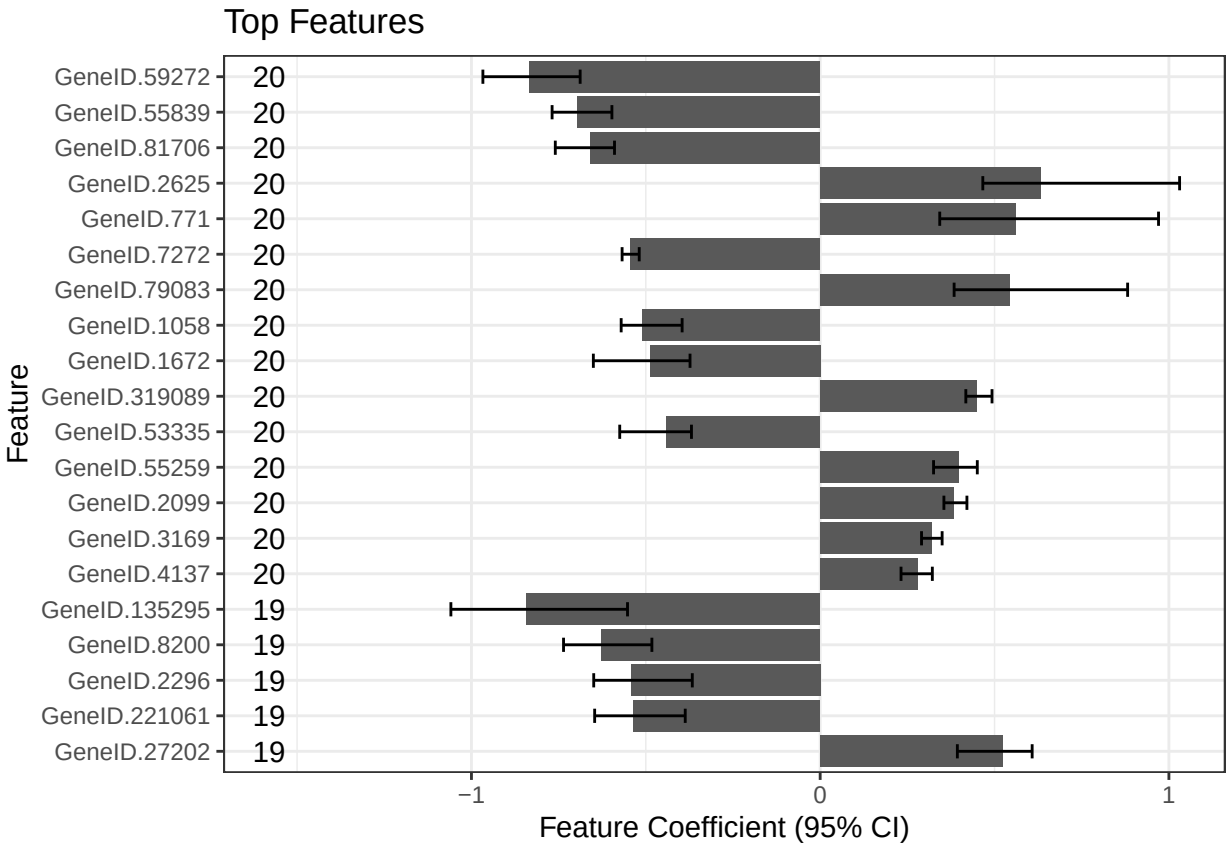
2. Important Features






```
## (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 5 CVs"

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   16.00  100.00  100.00   87.35  100.00  100.00
```


Top Features
GeneID.59272 20
GeneID.55839 20
GeneID.81706 20
GeneID.2625 20
GeneID.771 20
GeneID.7272 20
GeneID.79083 20
GeneID.1058 20
GeneID.1672 20
GeneID.319089 20
GeneID.53335 20
GeneID.55259 20
GeneID.2099 20
GeneID.3169 20
GeneID.4137 20
GeneID.135295 19
GeneID.8200 19
GeneID.2296 19
GeneID.221061 19
GeneID.27202 19
Feature
−1
0
1
Feature Coefficient (95% CI)

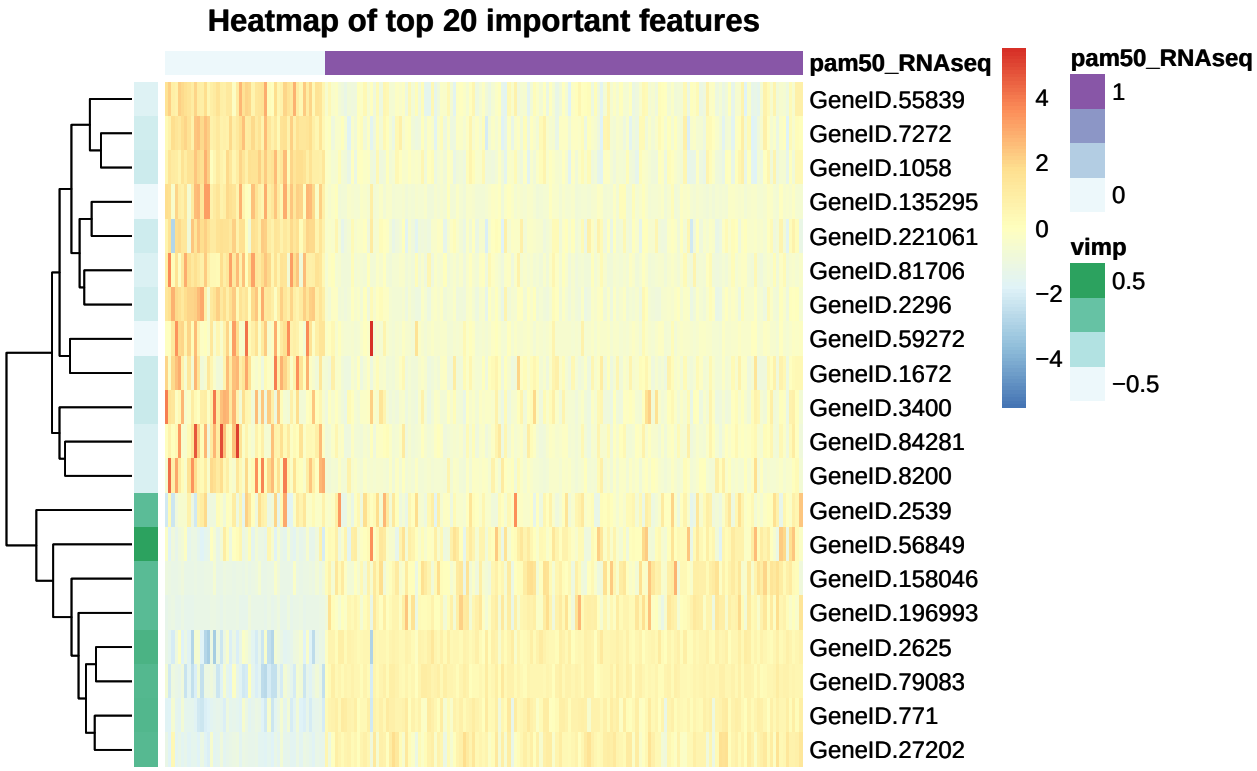

Heatmap of top 20 important features
pam50_RNAseq
GeneID.55839
GeneID.7272
GeneID.1058
GeneID.135295
GeneID.221061
GeneID.81706
GeneID.2296
GeneID.59272
GeneID.1672
GeneID.3400
GeneID.84281
GeneID.8200
GeneID.2539
GeneID.56849
GeneID.158046
GeneID.196993
GeneID.2625
GeneID.79083
GeneID.771
GeneID.27202
4
2
0
−2
−4
pam50_RNAseq
1
0
vimp
0.5
−0.5

Table 1: parameter selection

seed	alpha	lambda	cv
1001	0.0	0.0390268	1
1001	0.0	0.0393106	2
1001	0.0	0.0387182	3
1001	0.5	0.0005100	4
1001	0.0	0.0388060	5
1002	0.0	0.0388340	1
1002	0.0	0.0387347	2
1002	0.5	0.0005194	3
1002	0.0	0.0388192	4
1002	0.0	0.0389460	5
1003	0.0	0.0392710	1
1003	0.5	0.0005150	2
1003	0.0	0.0387054	3
1003	0.0	0.0389664	4
1003	0.0	0.0389099	5
1004	0.0	0.0387564	1
1004	1.0	0.0003695	2
1004	0.0	0.0391371	3
1004	0.0	0.0389803	4
1004	0.0	0.0391363	5