Evaluate testing data (multi-class) - Lasso

Andrew Chang

2018-12-05

Contents

0. Load Data . . . 1
1. Scores . . . 1
2. Important Features . . . 3

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

1. Lasso does not output 'size', so the "number of feature used"" to train the model is not informative. Should check 'lambda' which is correspond to the number of features used.

2. Lasso's vimp will be based on how many times the feature exist in all runs.

Labels: 0: Basal 1: LumA 2: LumB 3: Her2

```
## user input
project_home <- "~/EVE/tests"
project_name <- "lasso_multi_outCV_test"
```

0. Load Data

```
## 300 of samples were used
```

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

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

```
## Labels:
```

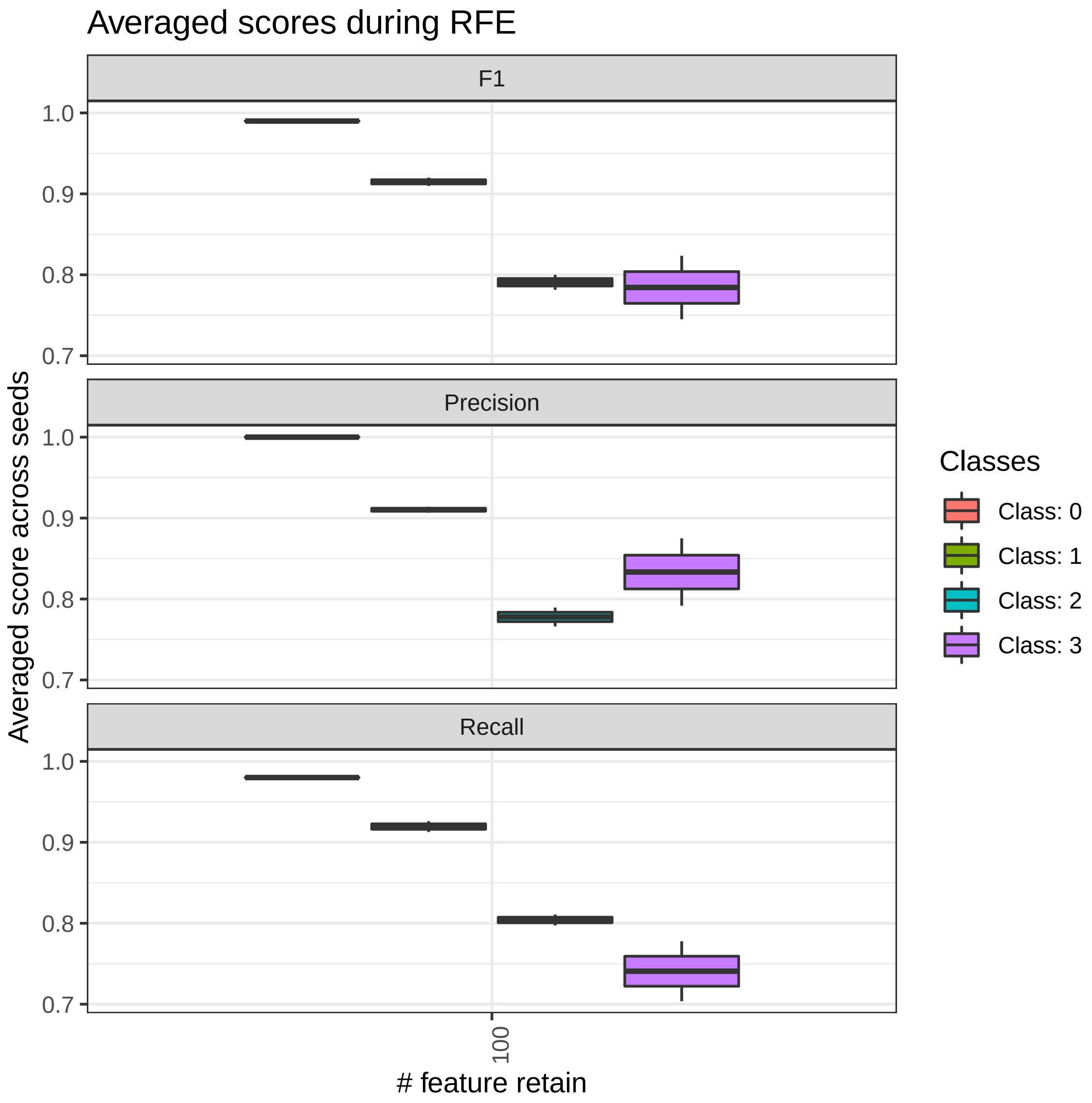
```
##
##   0   1   2   3
##  50 149  74  27
```

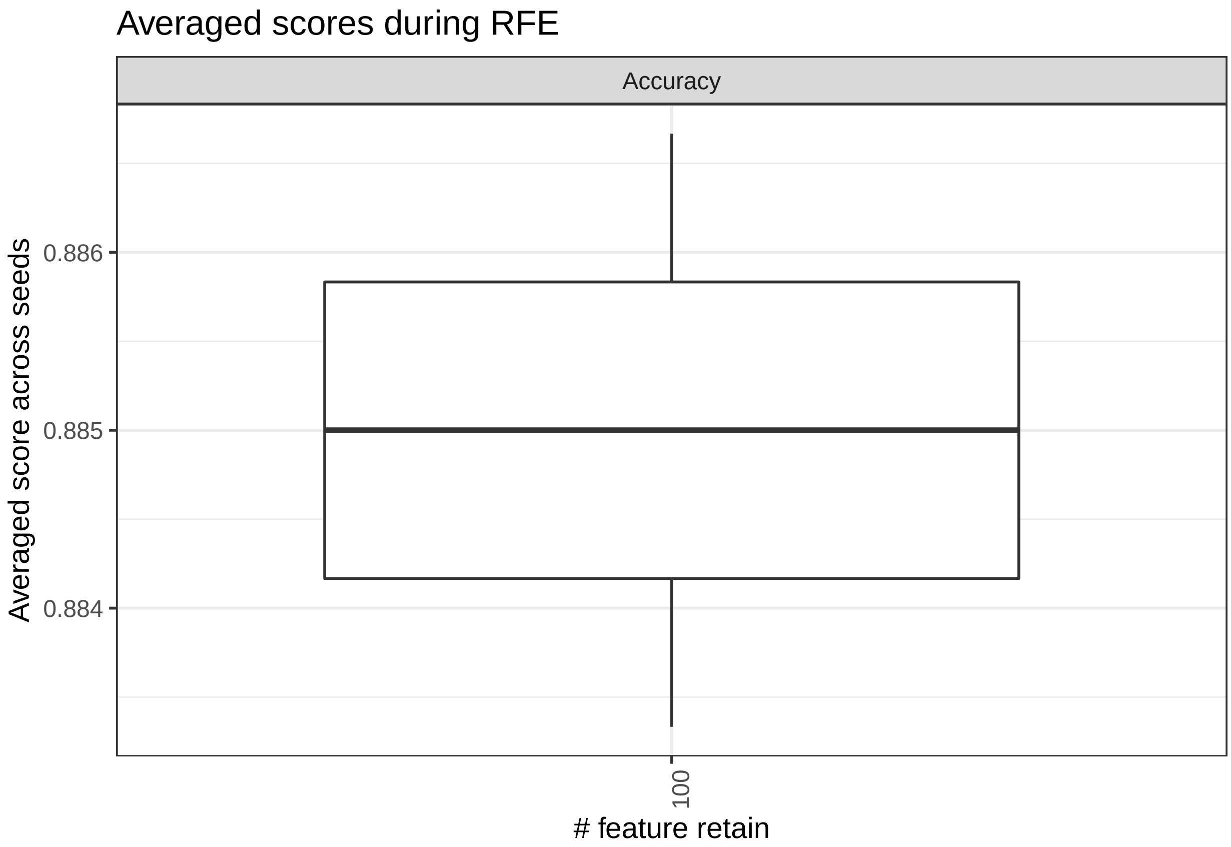
1. Scores

`average = T`: scores are based on pre-validation (combine predictions from all CVs per job and then calculate a single score per job).

`average = F`: report all the scores from each CV across entire jobs.

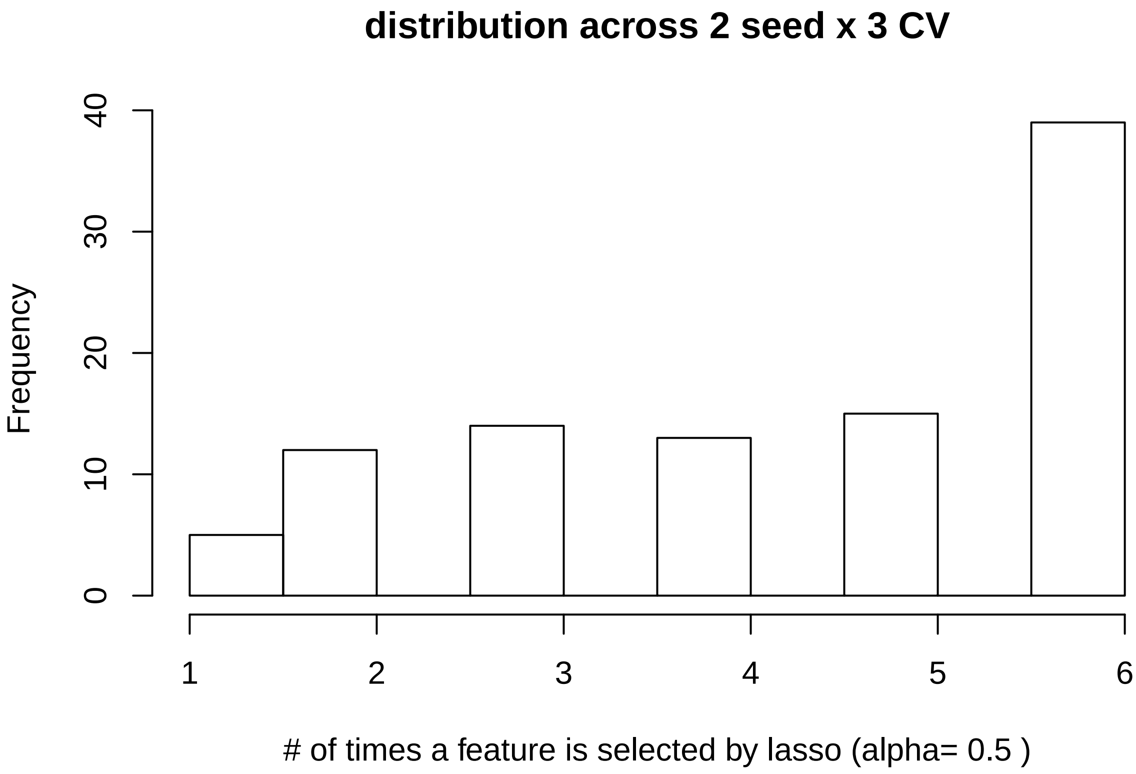
Note: '# of feature retain' is not meaningful in this setup. It is just the total number of features. The RFE concept is cooked into 'lambda' optimization, and each CV may have different lambda values, so it makes comparing prevalidation scores at different lambda difficult.

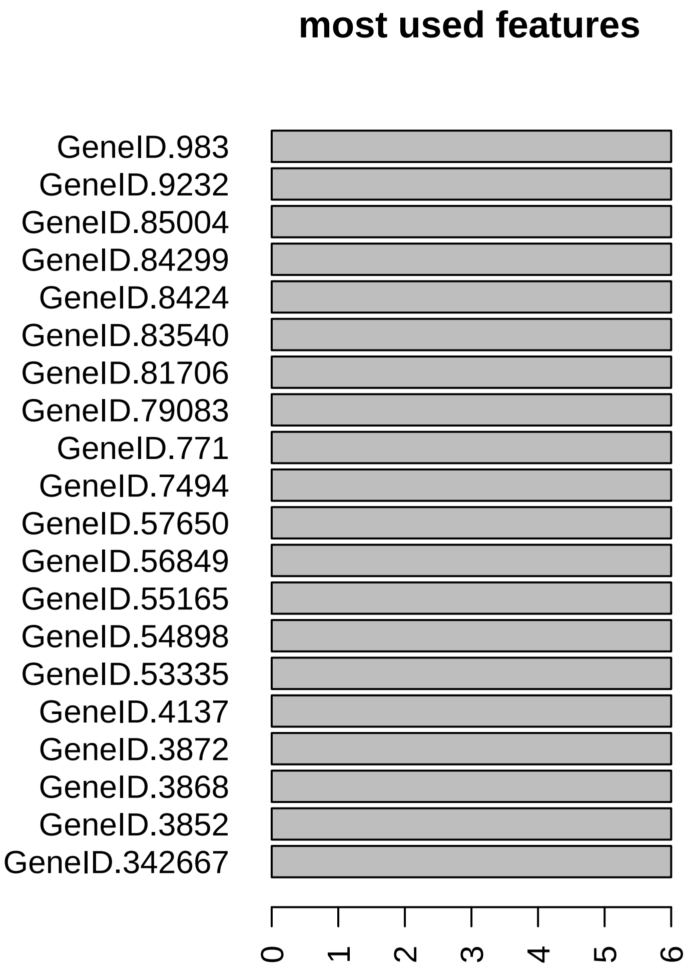

Averaged scores during RFE
F1
Precision
Recall
1.0
0.9
0.8
0.7
Averaged score across seeds
100
# feature retain
Classes
Class: 0
Class: 1
Class: 2
Class: 3



2. Important Features

For Lasso, we calculate how many times a given features is being used in all the runs.


distribution across 2 seed x 3 CV
Frequency
0
10
20
30
40
1
2
3
4
5
6
# of times a feature is selected by lasso (alpha= 0.5)


most used features
GeneID.983
GeneID.9232
GeneID.85004
GeneID.84299
GeneID.8424
GeneID.83540
GeneID.81706
GeneID.79083
GeneID.771
GeneID.7494
GeneID.57650
GeneID.56849
GeneID.55165
GeneID.54898
GeneID.53335
GeneID.4137
GeneID.3872
GeneID.3868
GeneID.3852
GeneID.342667
0
1
2
3
4
5
6