Evaluate testing data (binary-class) - xgboost

Andrew Chang

2018-12-05

Contents

Labels: 0: Basal 1: LumA

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

0. Load Data

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

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

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

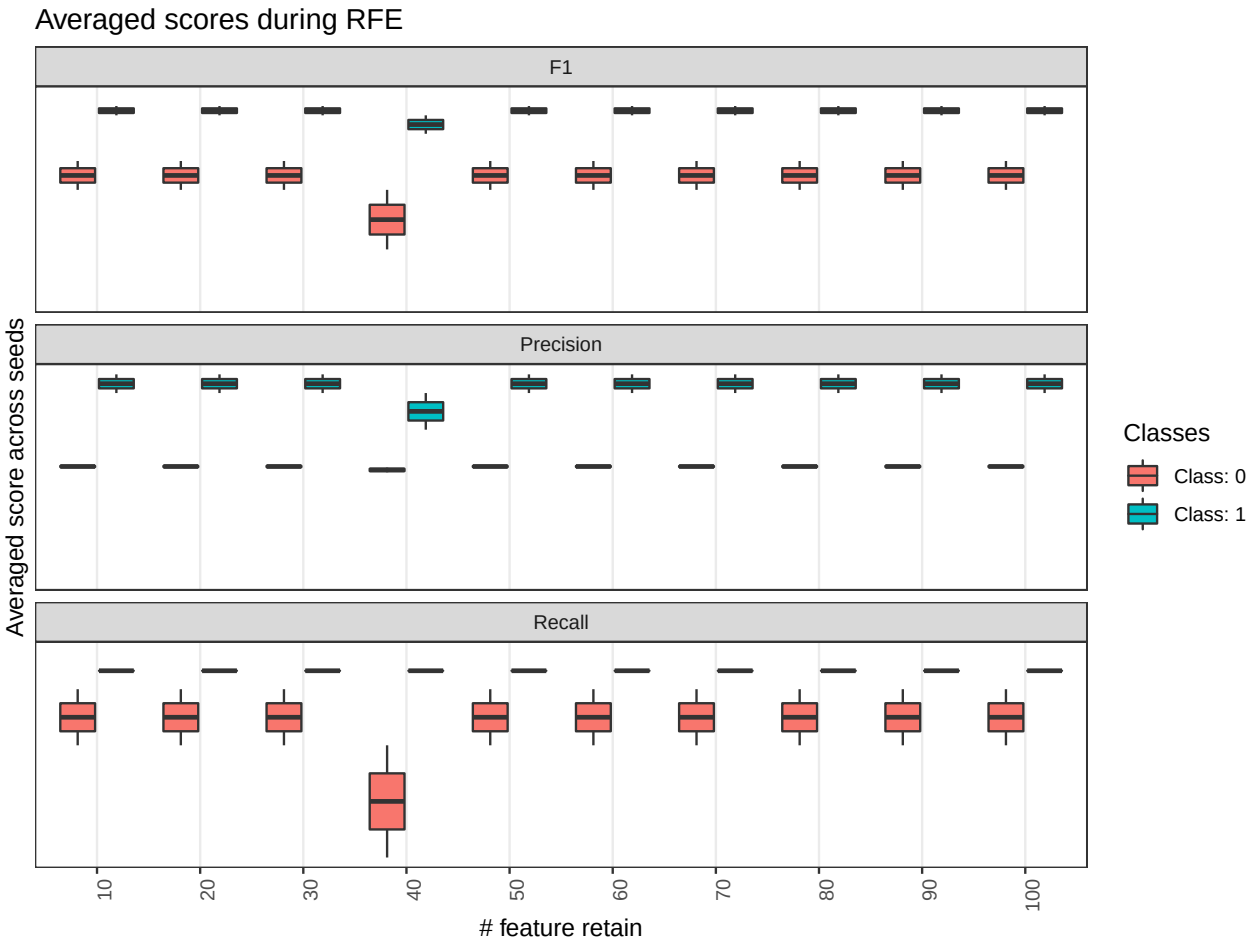
```
## Labels:
```

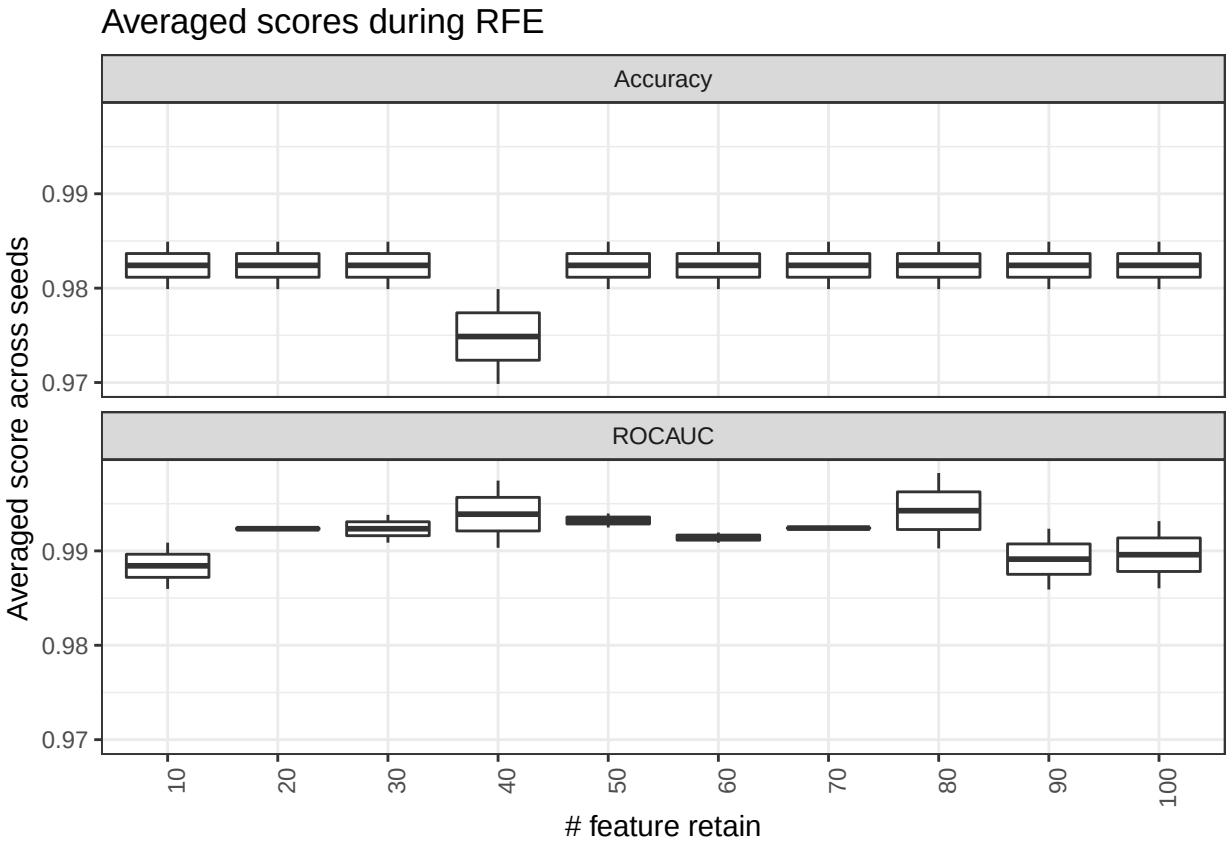
```
##
##   0   1
##  50 149
```

1. Scores

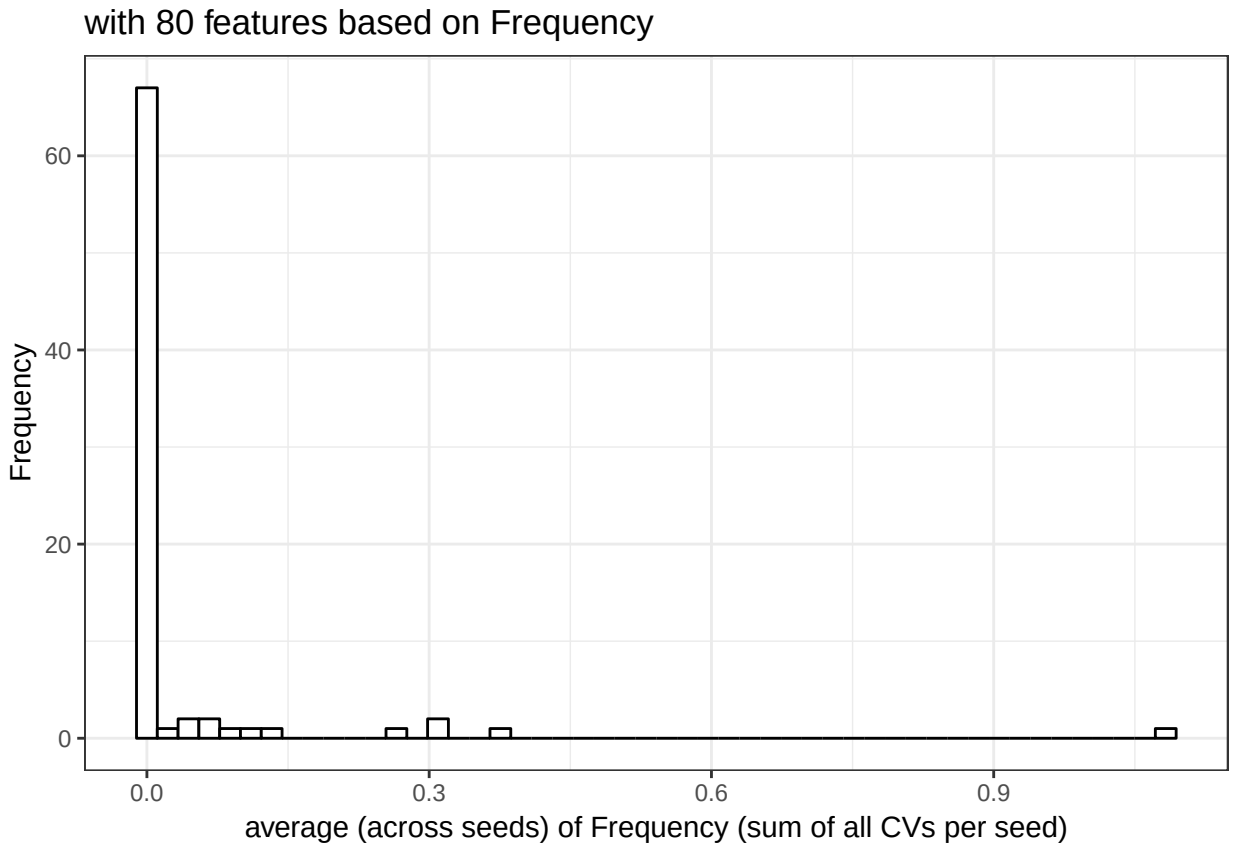
`average = T`: scores are based on pre-validation (combine predictions from all CVs per job and then calculate a single score per job).

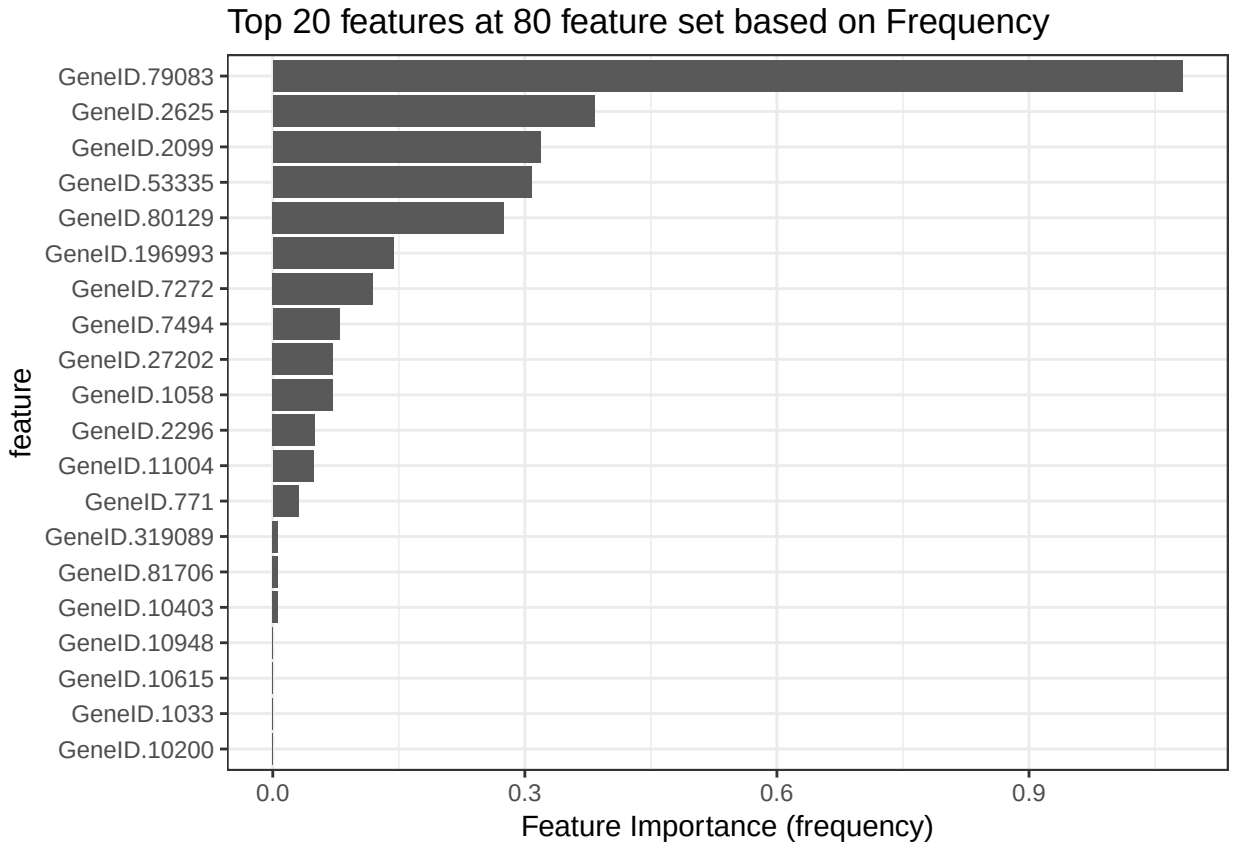
`average = F`: report all the scores from each CV across entire jobs.

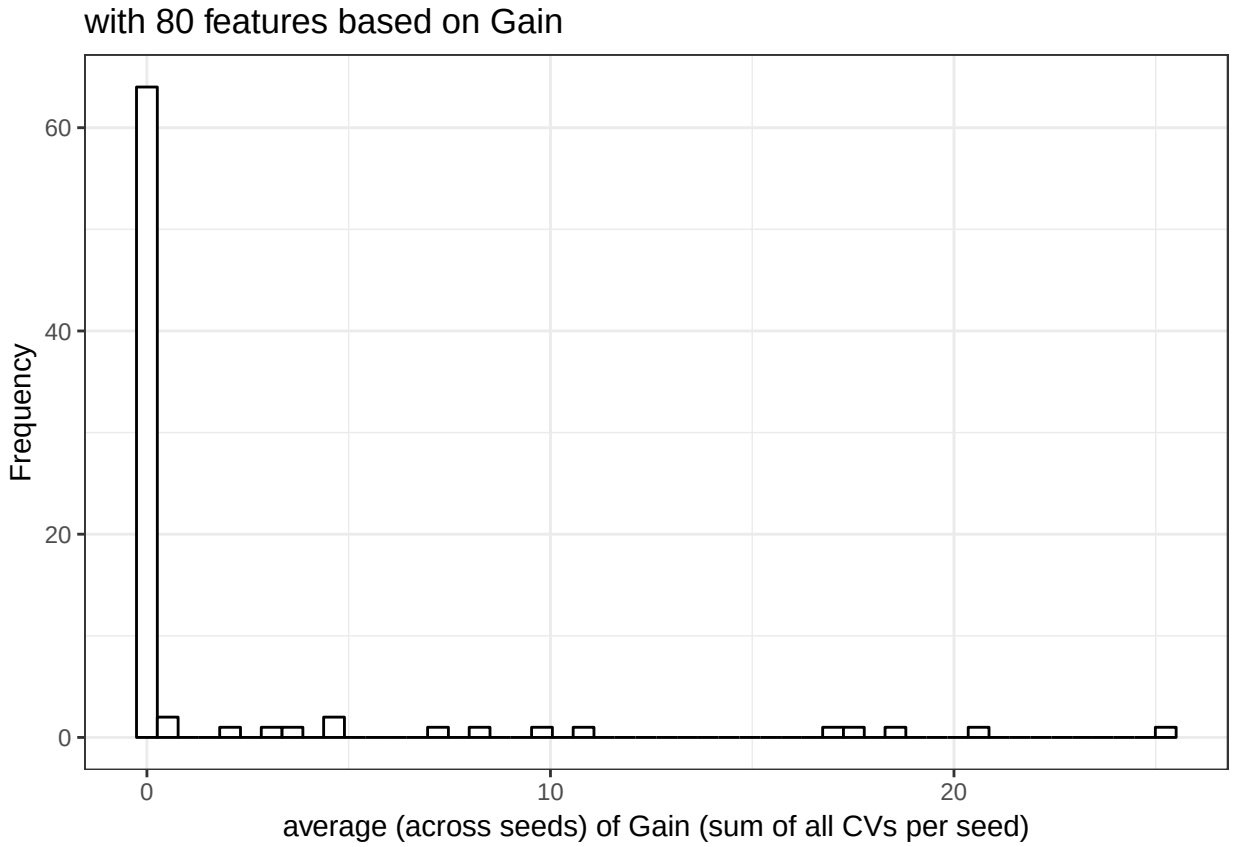

Averaged scores during RFE
F1
Precision
Recall
Averaged score across seeds
10
20
30
40
50
60
70
80
90
100
feature retain
Classes
Class: 0
Class: 1

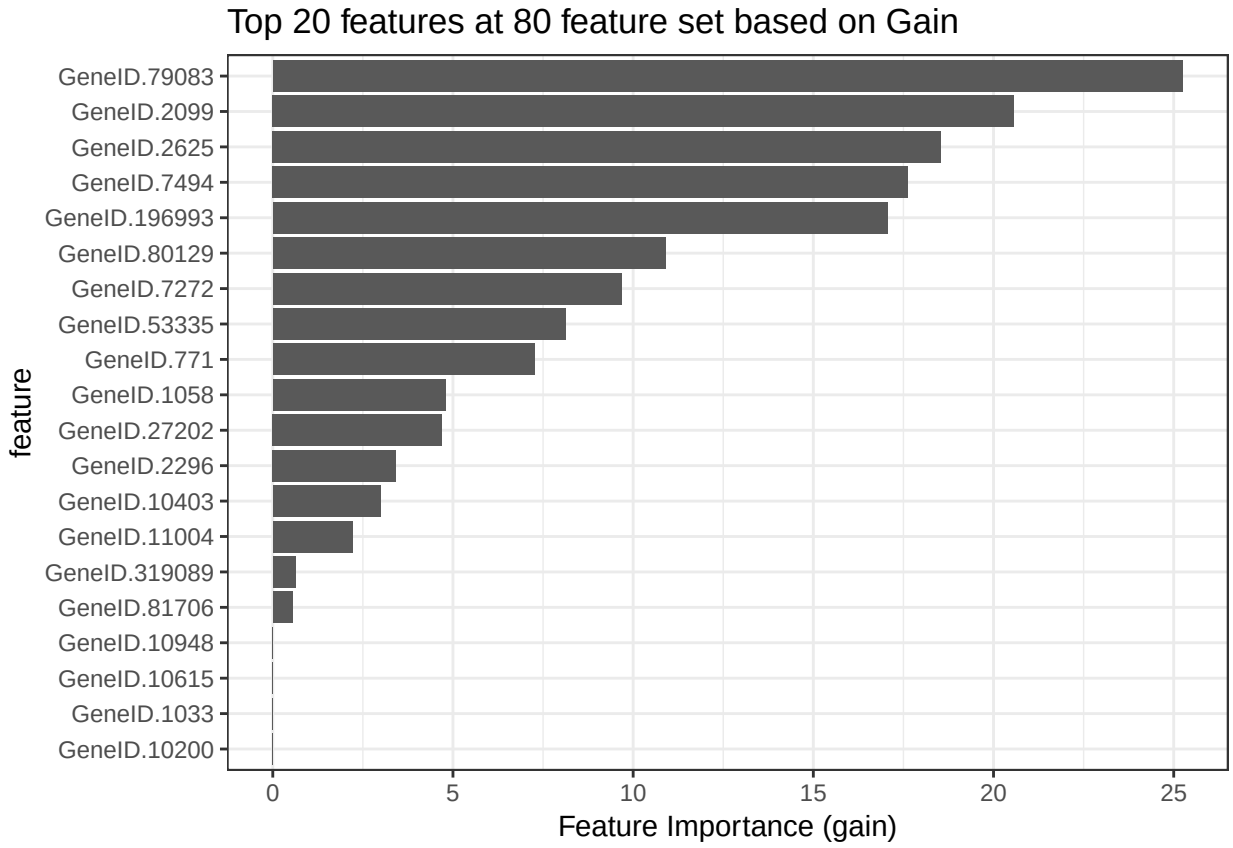

Averaged scores during RFE
Accuracy
ROCAUC
Averaged score across seeds
0.99
0.98
0.97
10
20
30
40
50
60
70
80
90
100
feature retain

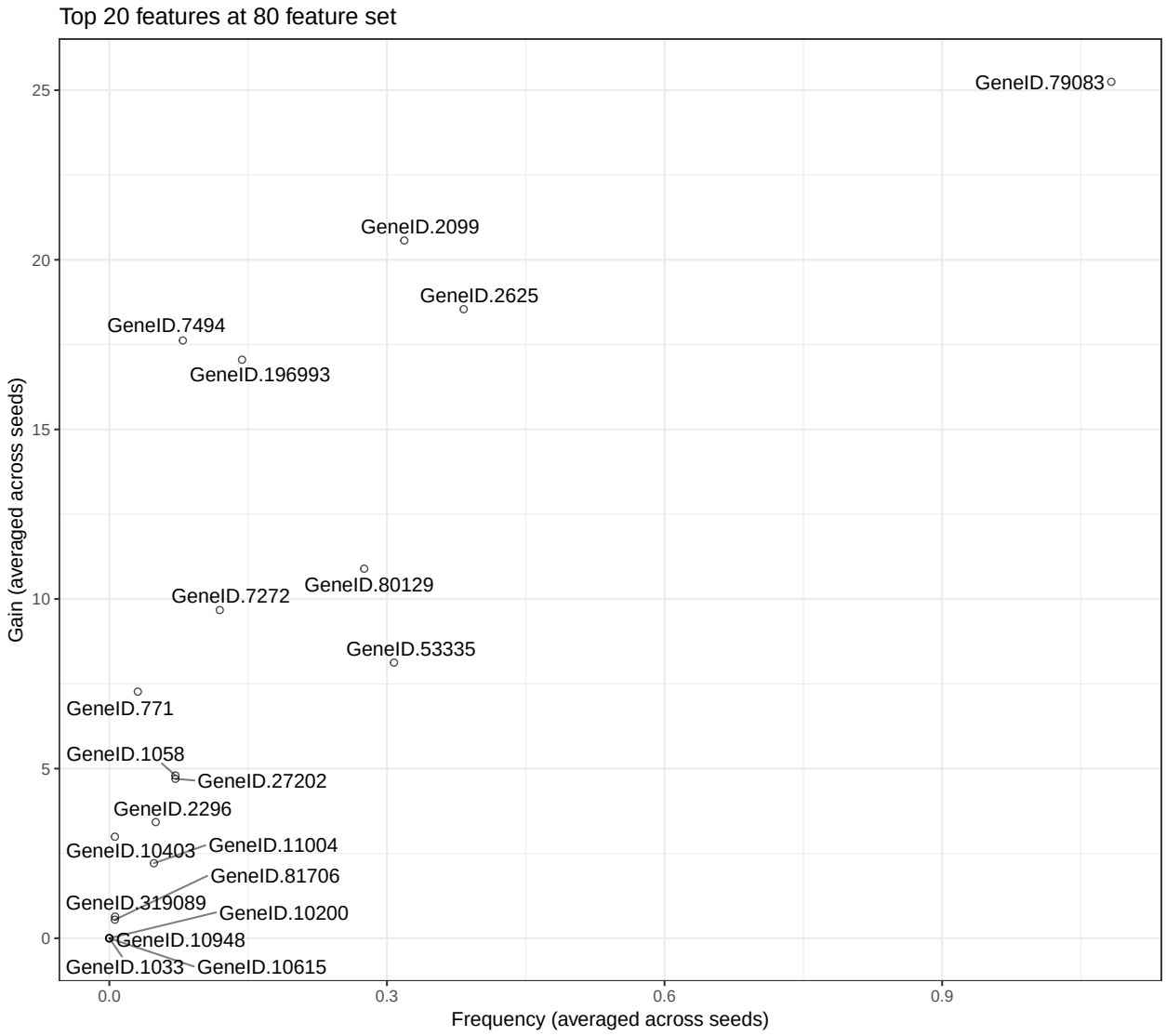
2. Important Features




Top 20 features at 80 feature set based on Frequency
GeneID.79083
GeneID.2625
GeneID.2099
GeneID.53335
GeneID.80129
GeneID.196993
GeneID.7272
GeneID.7494
GeneID.27202
GeneID.1058
GeneID.2296
GeneID.11004
GeneID.771
GeneID.319089
GeneID.81706
GeneID.10403
GeneID.10948
GeneID.10615
GeneID.1033
GeneID.10200
feature
0.0
0.3
0.6
0.9
Feature Importance (frequency)


with 80 features based on Gain
60
40
20
0
Frequency
0
10
20
average (across seeds) of Gain (sum of all CVs per seed)


Top 20 features at 80 feature set based on Gain
feature
GeneID.79083
GeneID.2099
GeneID.2625
GeneID.7494
GeneID.196993
GeneID.80129
GeneID.7272
GeneID.53335
GeneID.771
GeneID.1058
GeneID.27202
GeneID.2296
GeneID.10403
GeneID.11004
GeneID.319089
GeneID.81706
GeneID.10948
GeneID.10615
GeneID.1033
GeneID.10200
0
5
10
15
20
25
Feature Importance (gain)


Top 20 features at 80 feature set
Gain (averaged across seeds)
Frequency (averaged across seeds)
25
20
15
10
5
0
0.0
0.3
0.6
0.9
GeneID.79083
GeneID.2099
GeneID.2625
GeneID.7494
GeneID.196993
GeneID.80129
GeneID.7272
GeneID.53335
GeneID.771
GeneID.1058
GeneID.27202
GeneID.2296
GeneID.10403
GeneID.11004
GeneID.81706
GeneID.319089
GeneID.10200
GeneID.10948
GeneID.1033
GeneID.10615

3. Hyper-parameters