

Evaluate testing data (multi-class) - xgboost

Andrew Chang

2018-12-04

Contents

0. Load Data	1
1. Scores	1
2. Important Features	4
3. Hyper-parameters	8

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

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

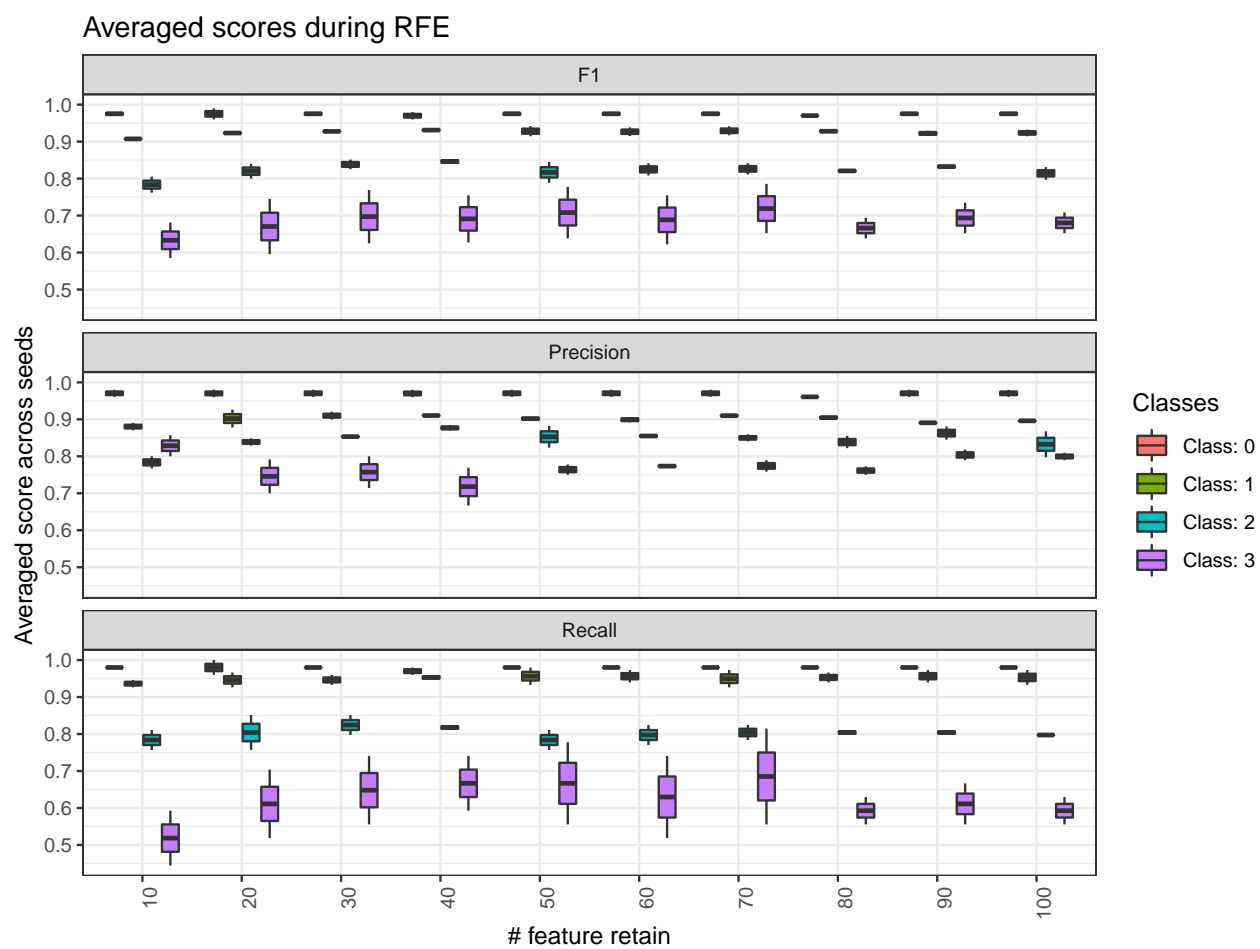
0. Load Data

```
## 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
```

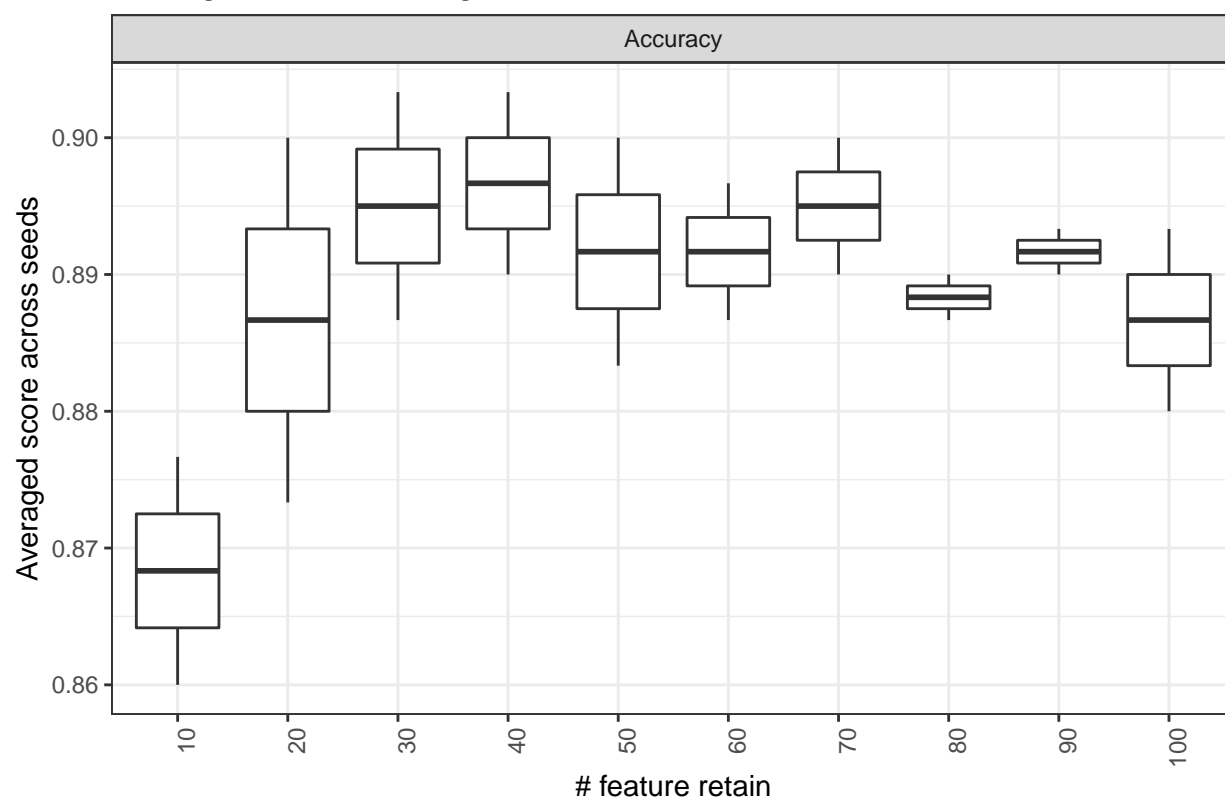
1. Scores

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.

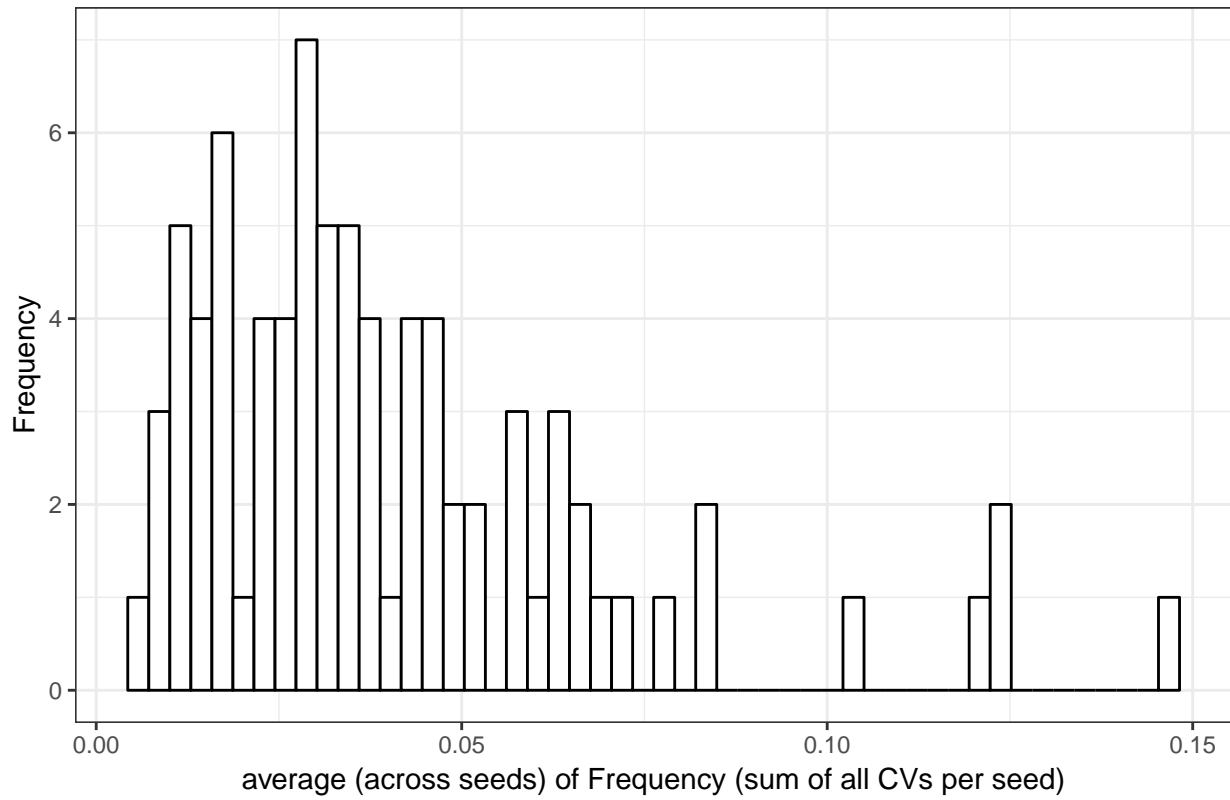


Averaged scores during RFE

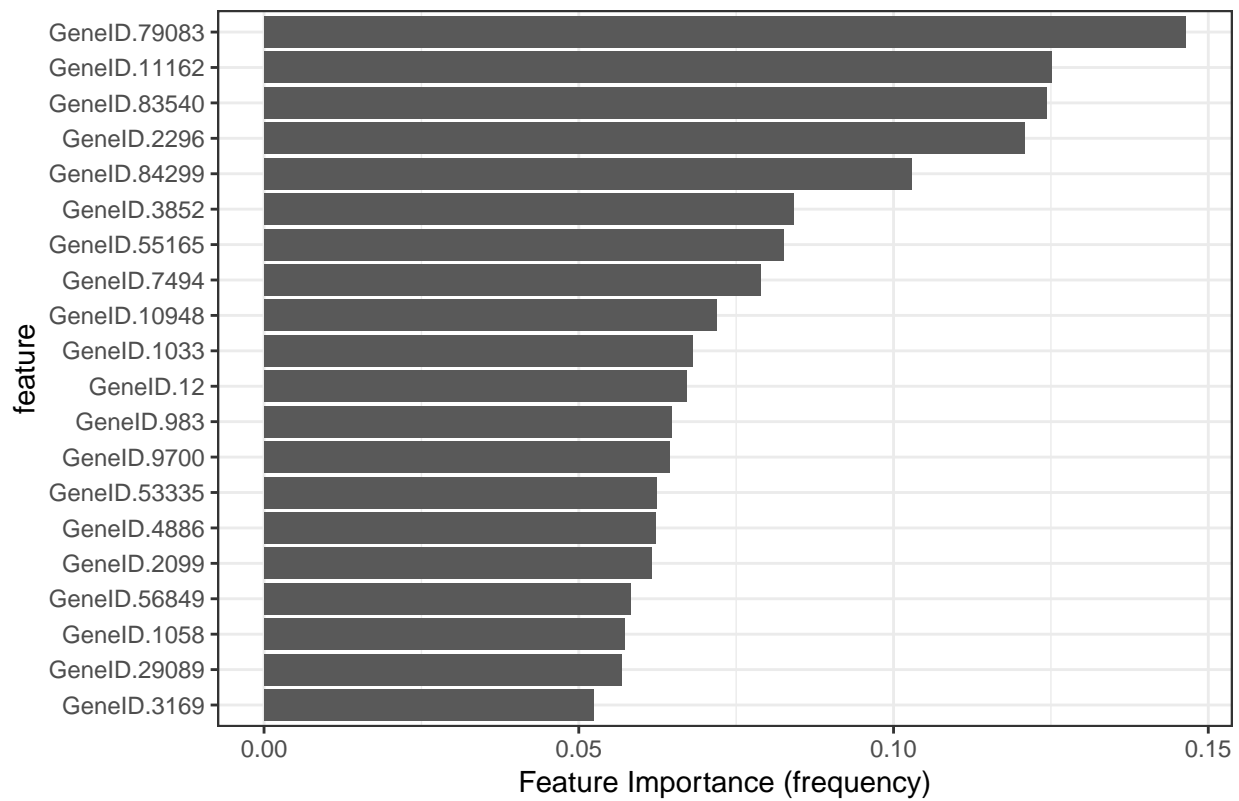


2. Important Features

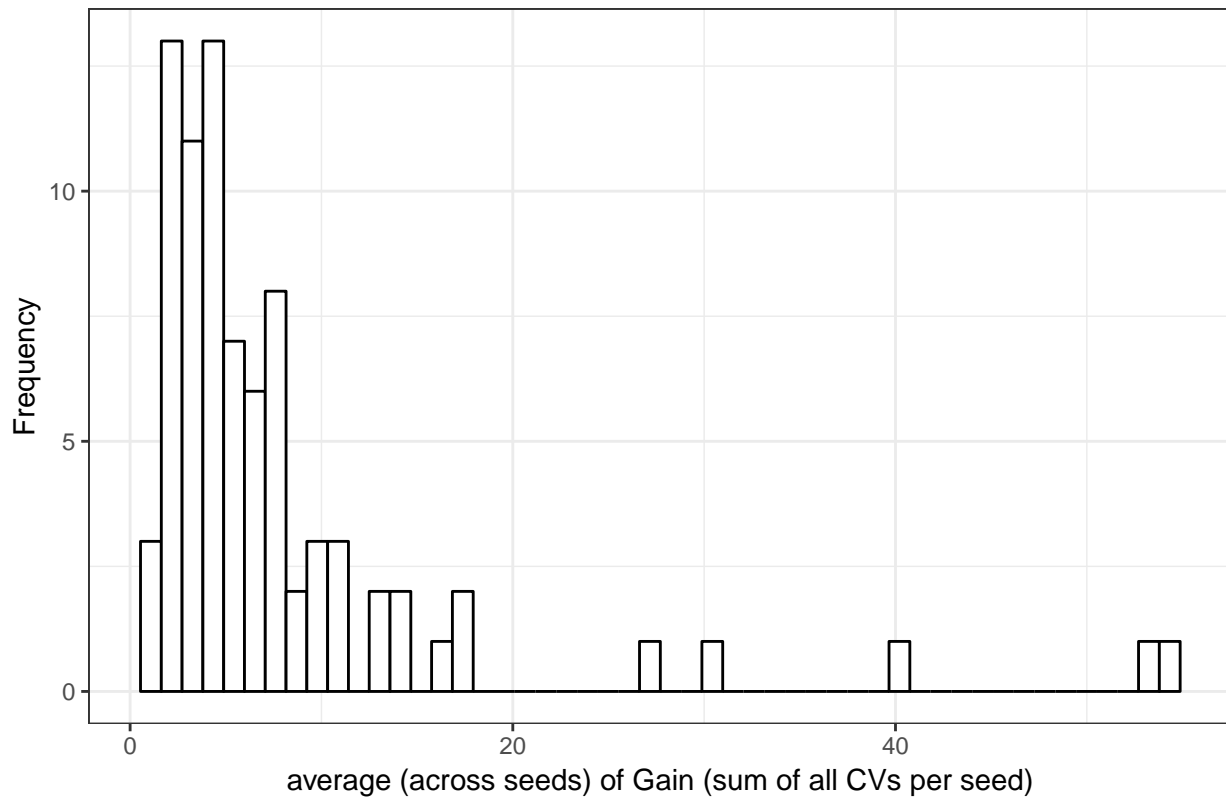
with 40 features based on Frequency



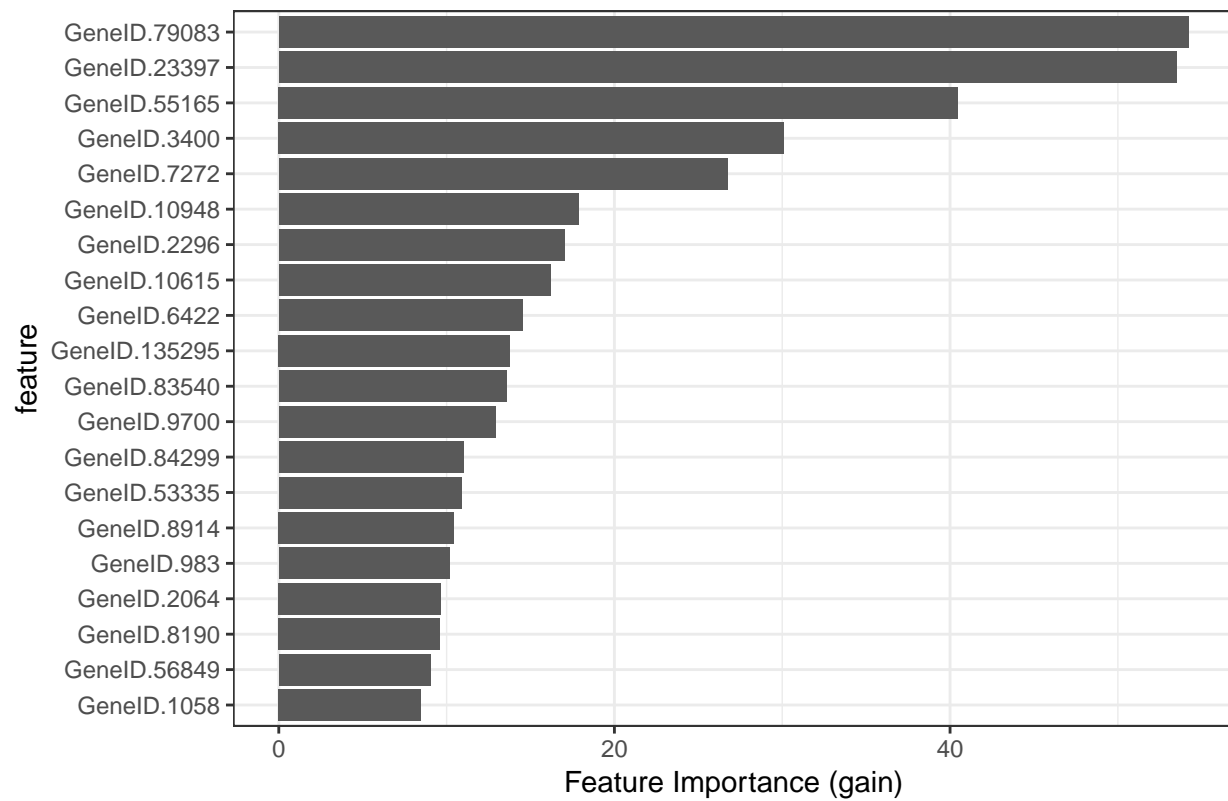
Top 20 features at 40 feature set based on Frequency



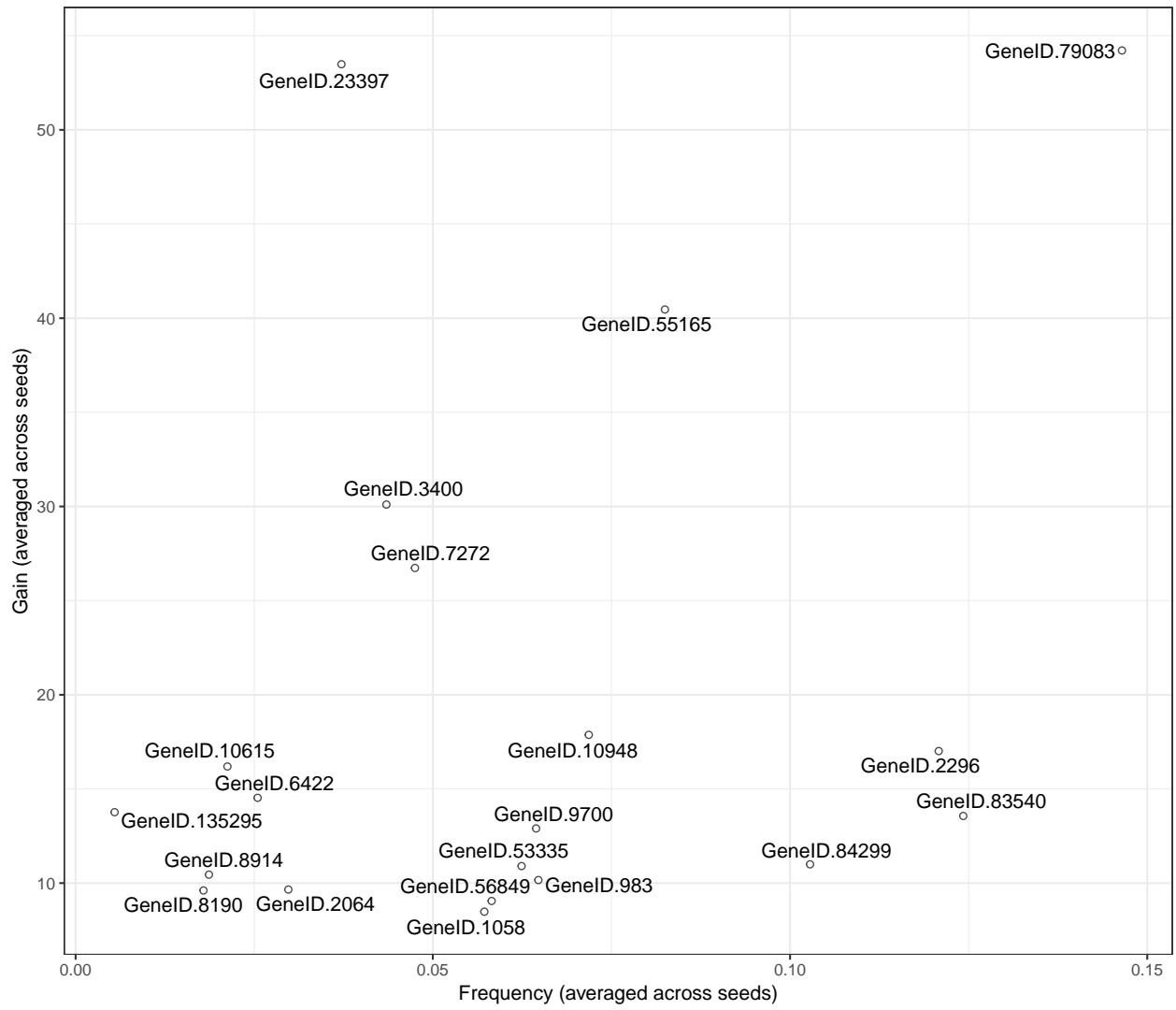
with 40 features based on Gain



Top 20 features at 40 feature set based on Gain



Top 20 features at 40 feature set



3. Hyper-parameters

