Evaluate testing data (survival) - XGBoost

John Doe

2019-01-30

Contents

Label: os_time

```
## user input
project_home <- "~/EVE/examples"
project_name <- "xgboost_survival_outCV_test"
```

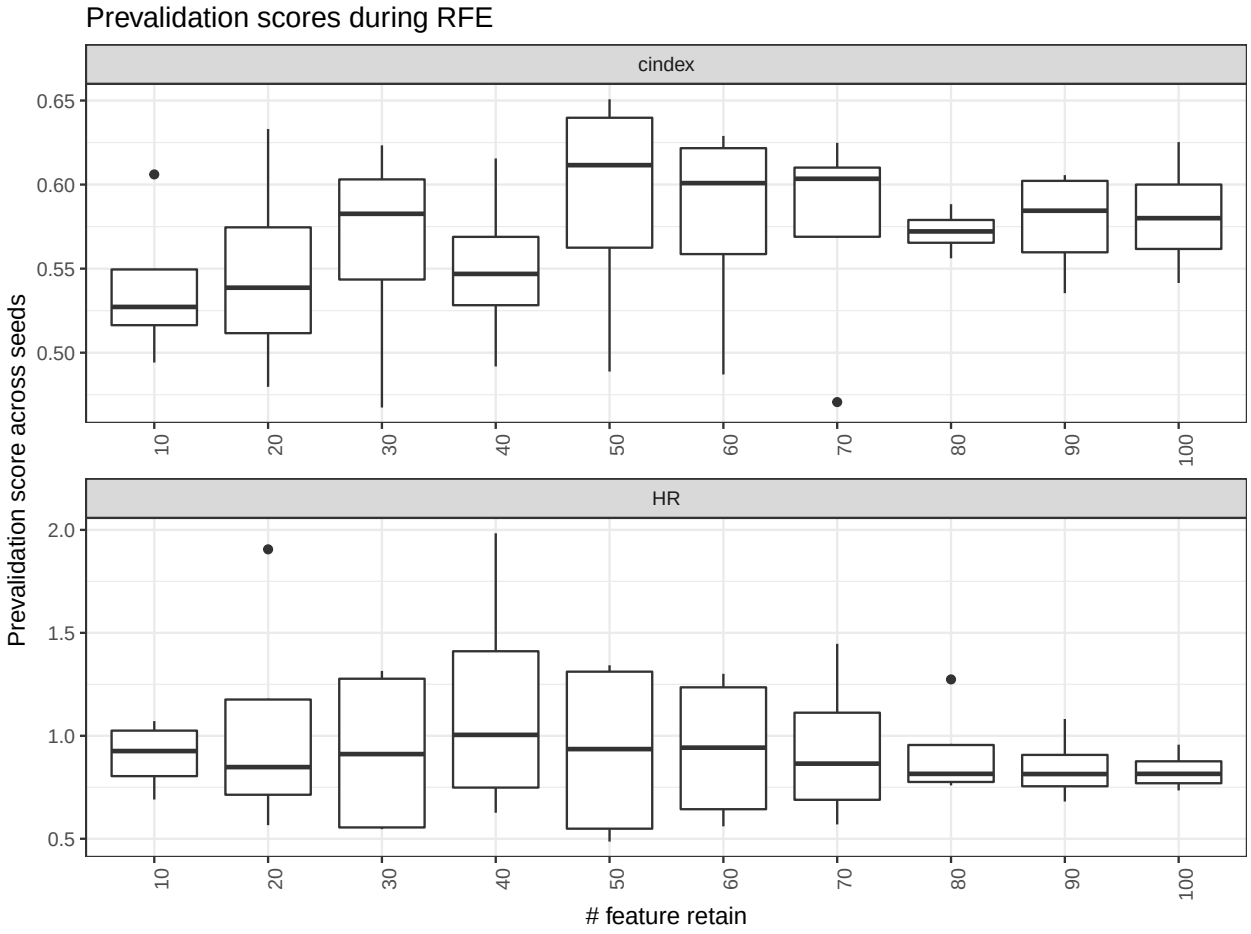
0. Load Data

```
300 of samples were used
```

```
100 of full features
```

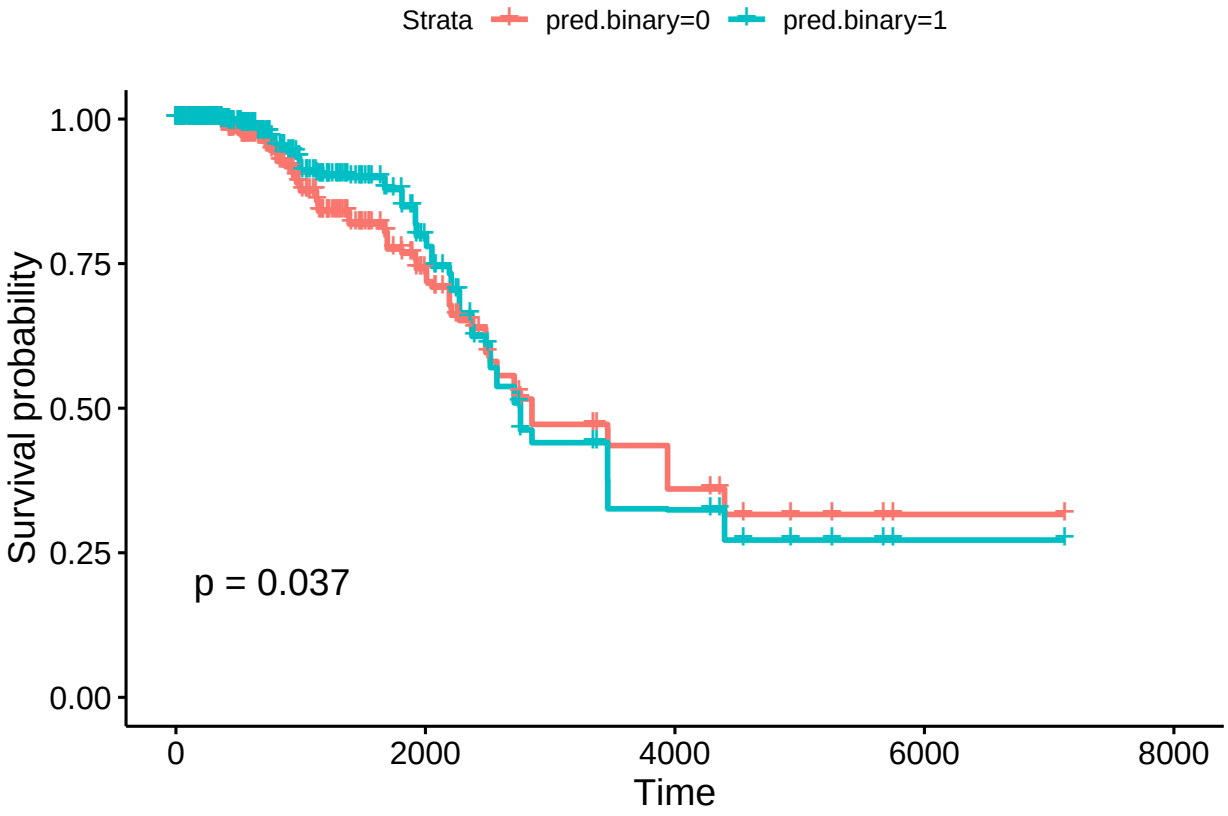
```
4 runs, each run contains 3 CVs.
```

1. Scores

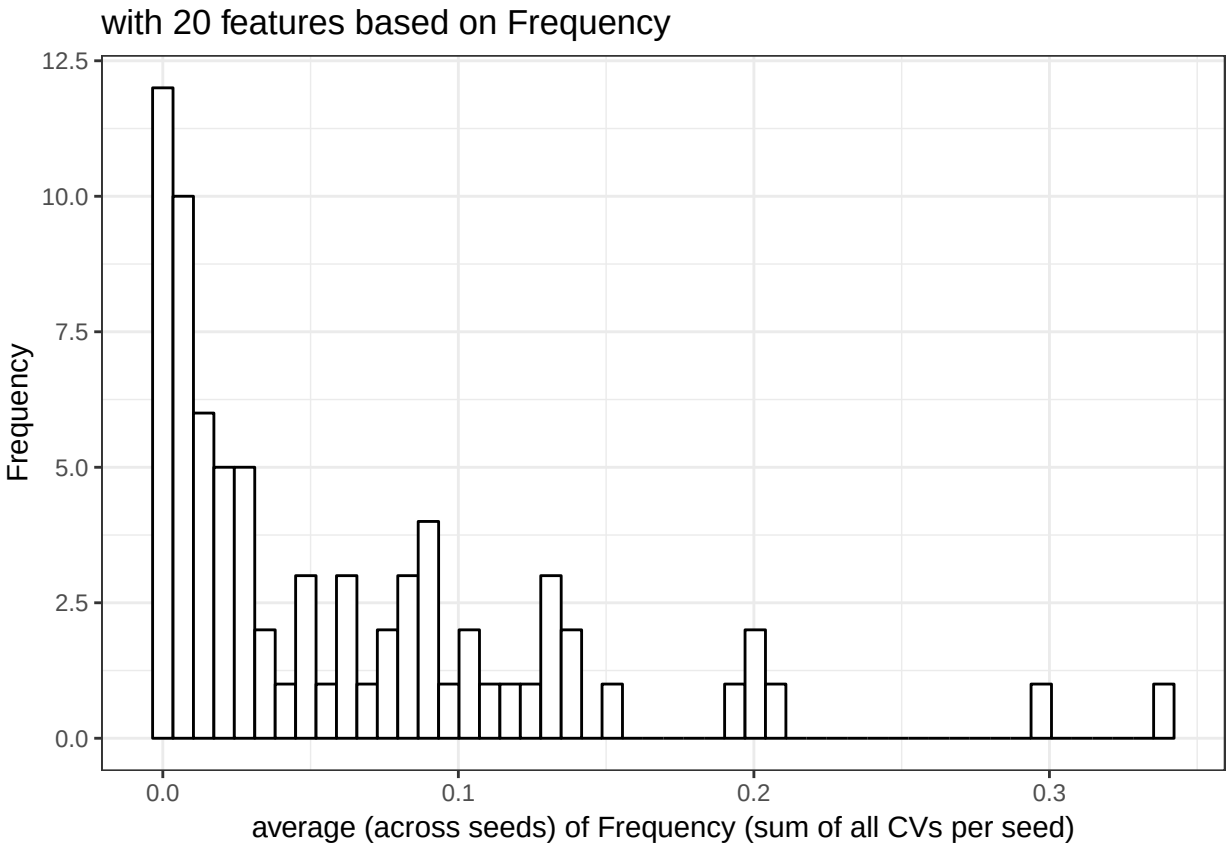


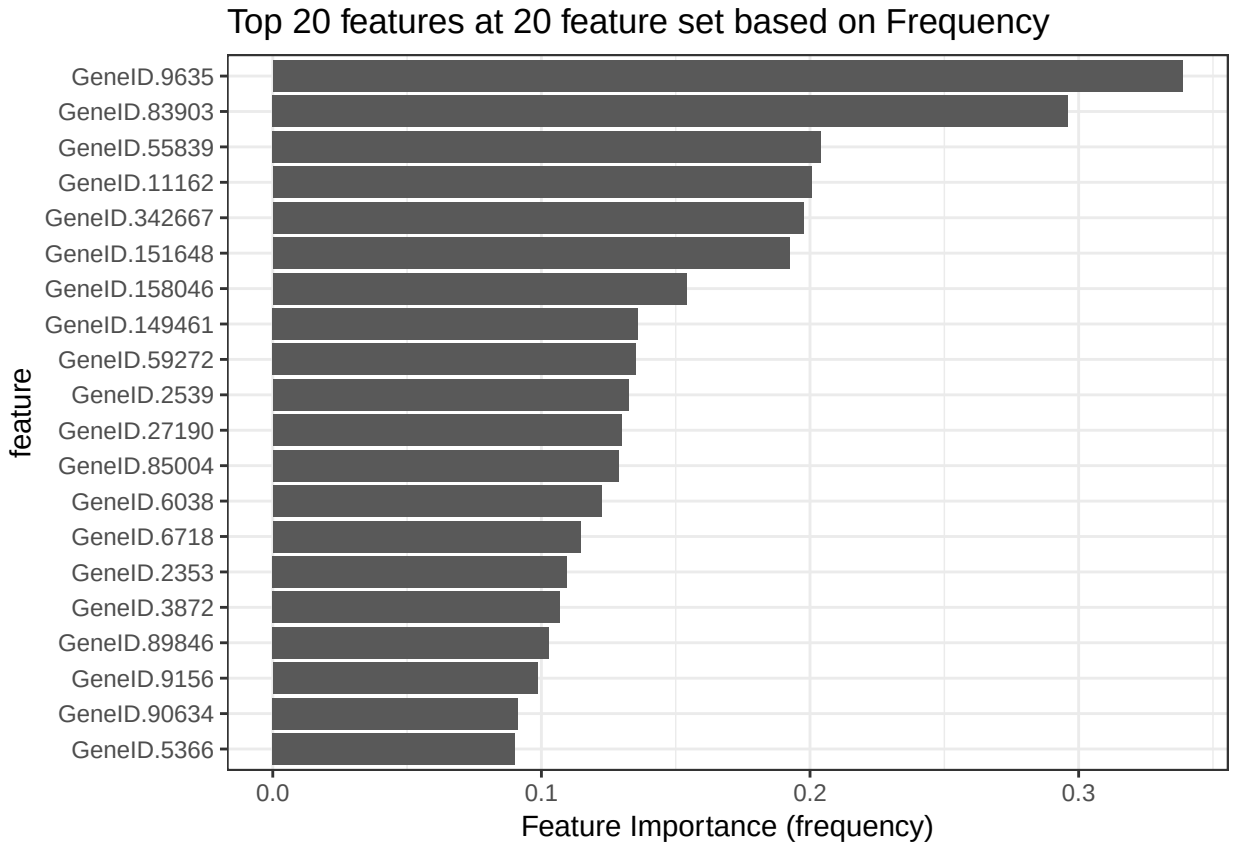
metrics	size.max	median.max	size.min	median.min
cindex	50	0.612	10	0.527
HR	40	1.004	90	0.814

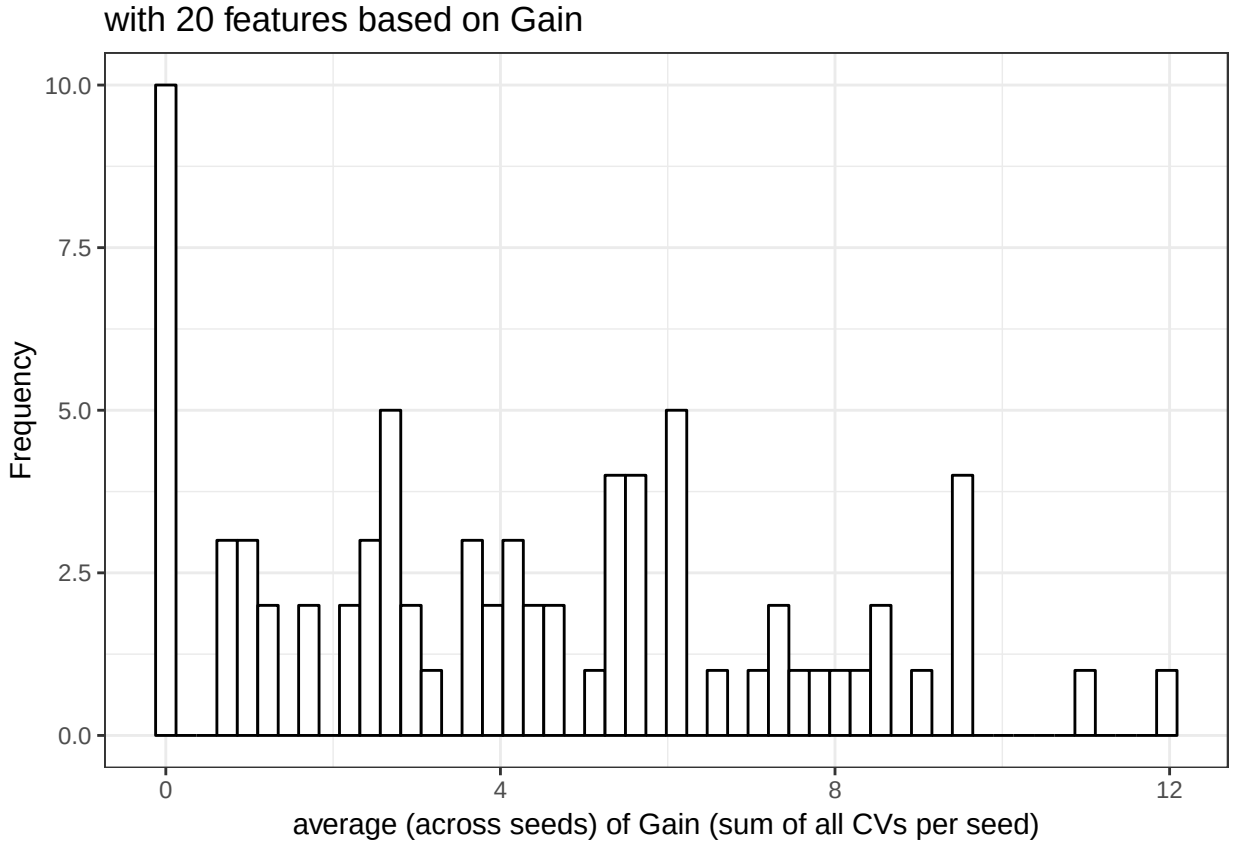
The following plot is to quickly see how well the prediction can separate long and short survivor.

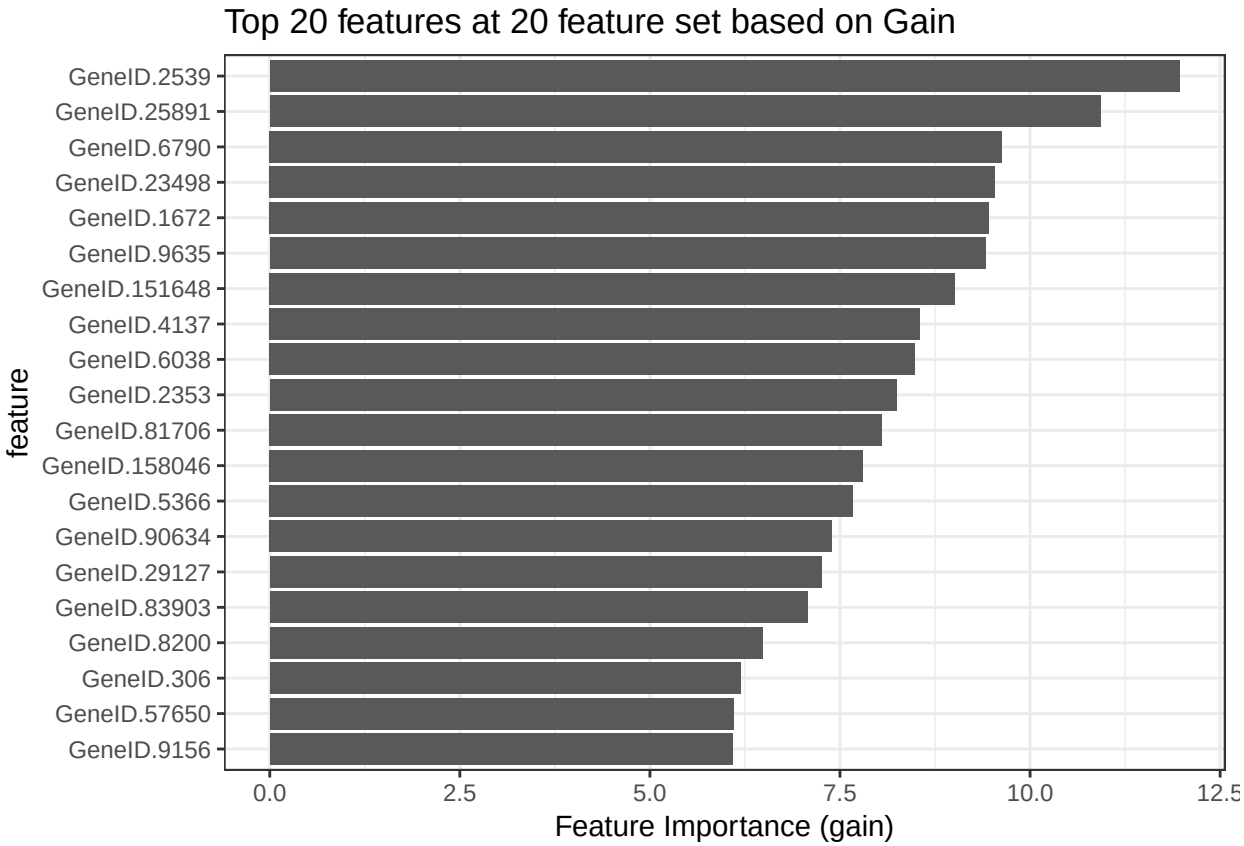

Strata
pred.binary=0
pred.binary=1
Survival probability
1.00
0.75
0.50
0.25
0.00
p = 0.037
0
2000
4000
6000
8000
Time

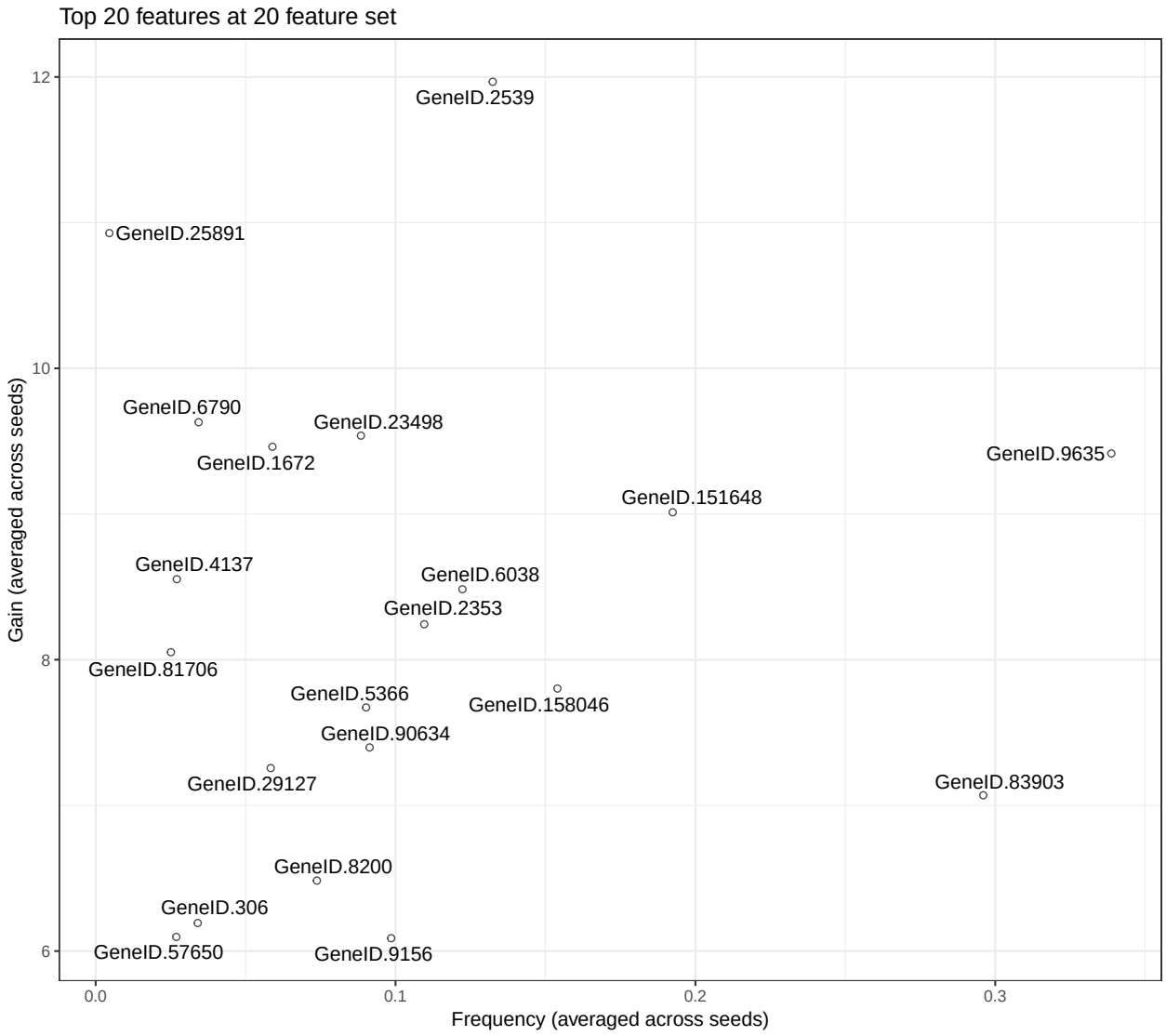
2. Important Features

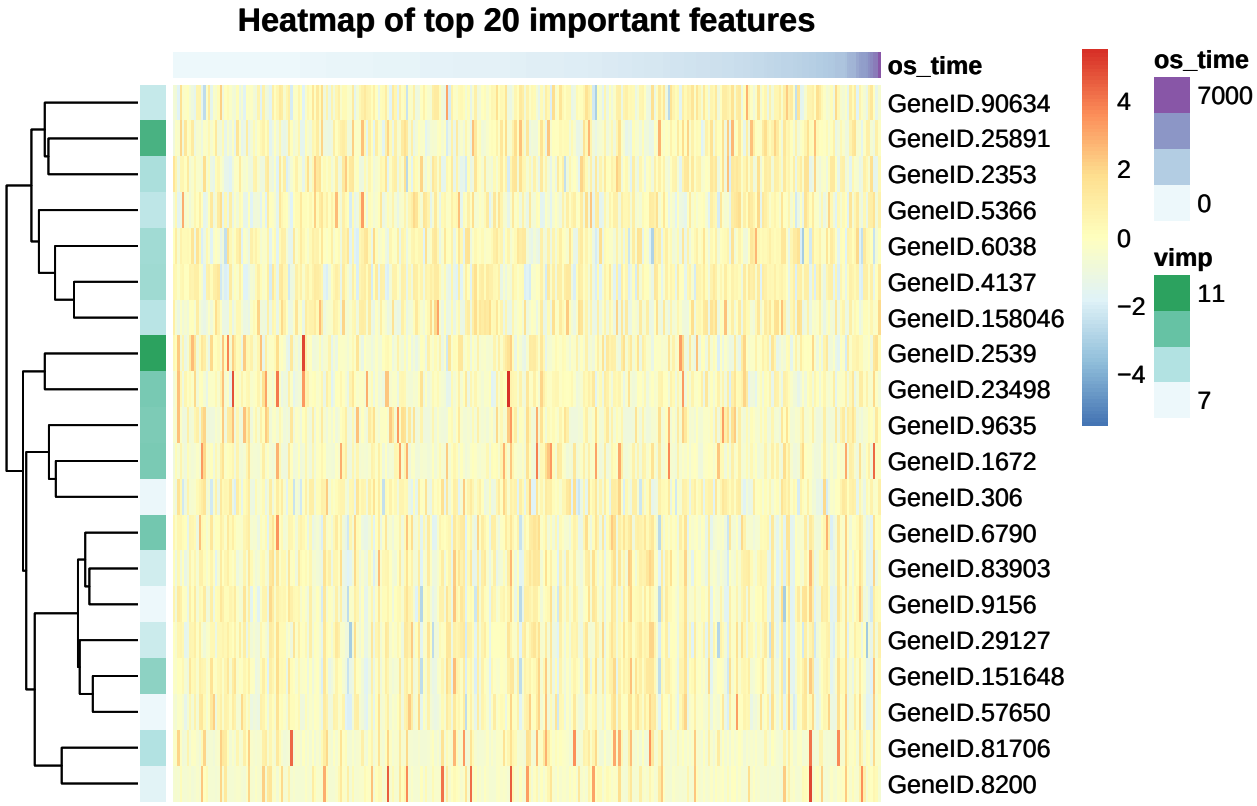



Top 20 features at 20 feature set based on Frequency
GeneID.9635
GeneID.83903
GeneID.55839
GeneID.11162
GeneID.342667
GeneID.151648
GeneID.158046
GeneID.149461
GeneID.59272
GeneID.2539
GeneID.27190
GeneID.85004
GeneID.6038
GeneID.6718
GeneID.2353
GeneID.3872
GeneID.89846
GeneID.9156
GeneID.90634
GeneID.5366
feature
0.0
0.1
0.2
0.3
Feature Importance (frequency)


with 20 features based on Gain
10.0
7.5
5.0
2.5
0.0
Frequency
0
4
8
12
average (across seeds) of Gain (sum of all CVs per seed)


Top 20 features at 20 feature set based on Gain
feature
GeneID.2539
GeneID.25891
GeneID.6790
GeneID.23498
GeneID.1672
GeneID.9635
GeneID.151648
GeneID.4137
GeneID.6038
GeneID.2353
GeneID.81706
GeneID.158046
GeneID.5366
GeneID.90634
GeneID.29127
GeneID.83903
GeneID.8200
GeneID.306
GeneID.57650
GeneID.9156
0.0
2.5
5.0
7.5
10.0
12.5
Feature Importance (gain)


Top 20 features at 20 feature set
Gain (averaged across seeds)
Frequency (averaged across seeds)
12
10
8
6
0.0
0.1
0.2
0.3
GeneID.2539
GeneID.25891
GeneID.6790
GeneID.23498
GeneID.1672
GeneID.9635
GeneID.151648
GeneID.4137
GeneID.6038
GeneID.2353
GeneID.81706
GeneID.5366
GeneID.158046
GeneID.90634
GeneID.29127
GeneID.83903
GeneID.8200
GeneID.306
GeneID.57650
GeneID.9156


Heatmap of top 20 important features
os_time
GeneID.90634
GeneID.25891
GeneID.2353
GeneID.5366
GeneID.6038
GeneID.4137
GeneID.158046
GeneID.2539
GeneID.23498
GeneID.9635
GeneID.1672
GeneID.306
GeneID.6790
GeneID.83903
GeneID.9156
GeneID.29127
GeneID.151648
GeneID.57650
GeneID.81706
GeneID.8200
4
2
0
−2
−4
os_time
7000
0
vimp
11
7

3. Hyper-parameters