Evaluate testing data (regression) - xgboost

Andrew Chang

2018-12-04

Contents

Labels: os_time

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

## 0. Load Data

```
## 300 of samples were used

## 101 of full features

## 2 runs, each run contains 3 CVs.

## os_time :

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     0.0   182.8   480.0   889.4  1221.2  7125.0
```
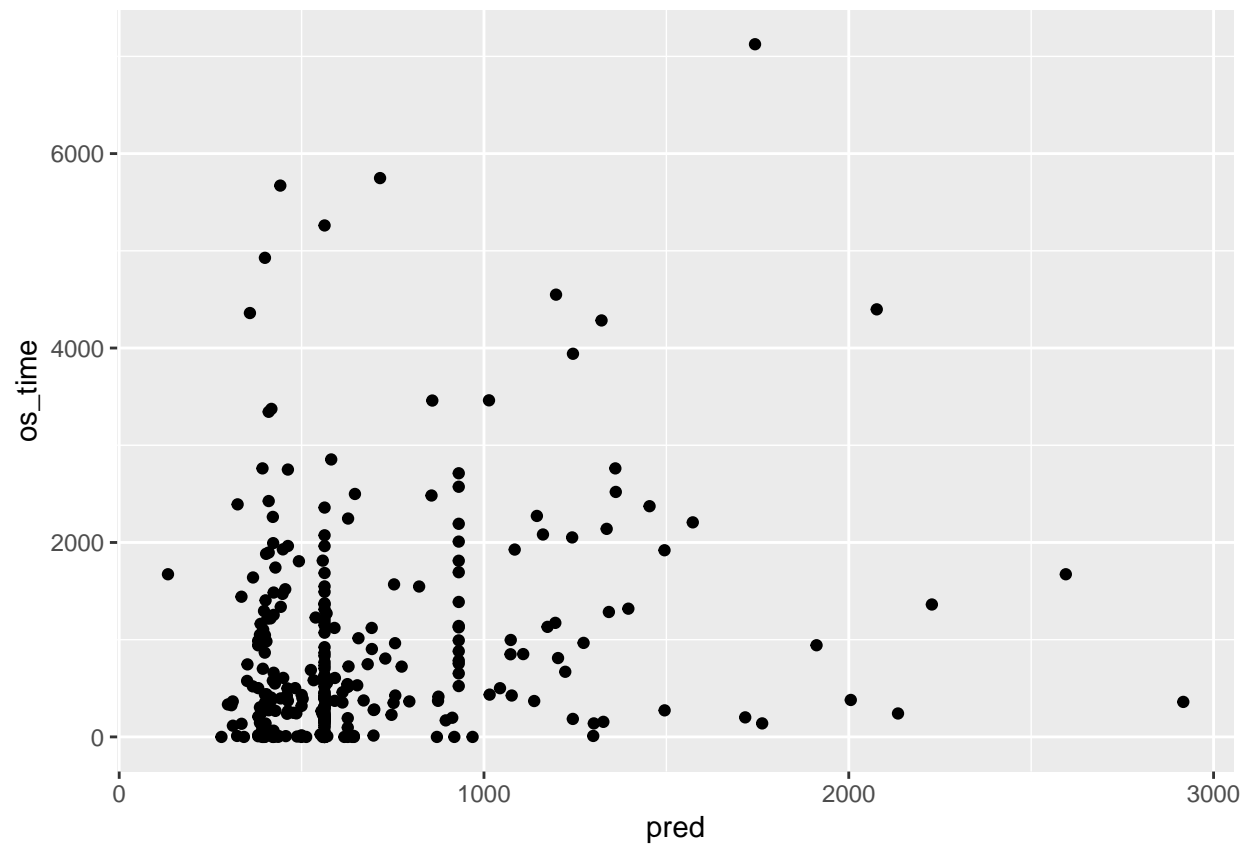
1. Scores

`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.
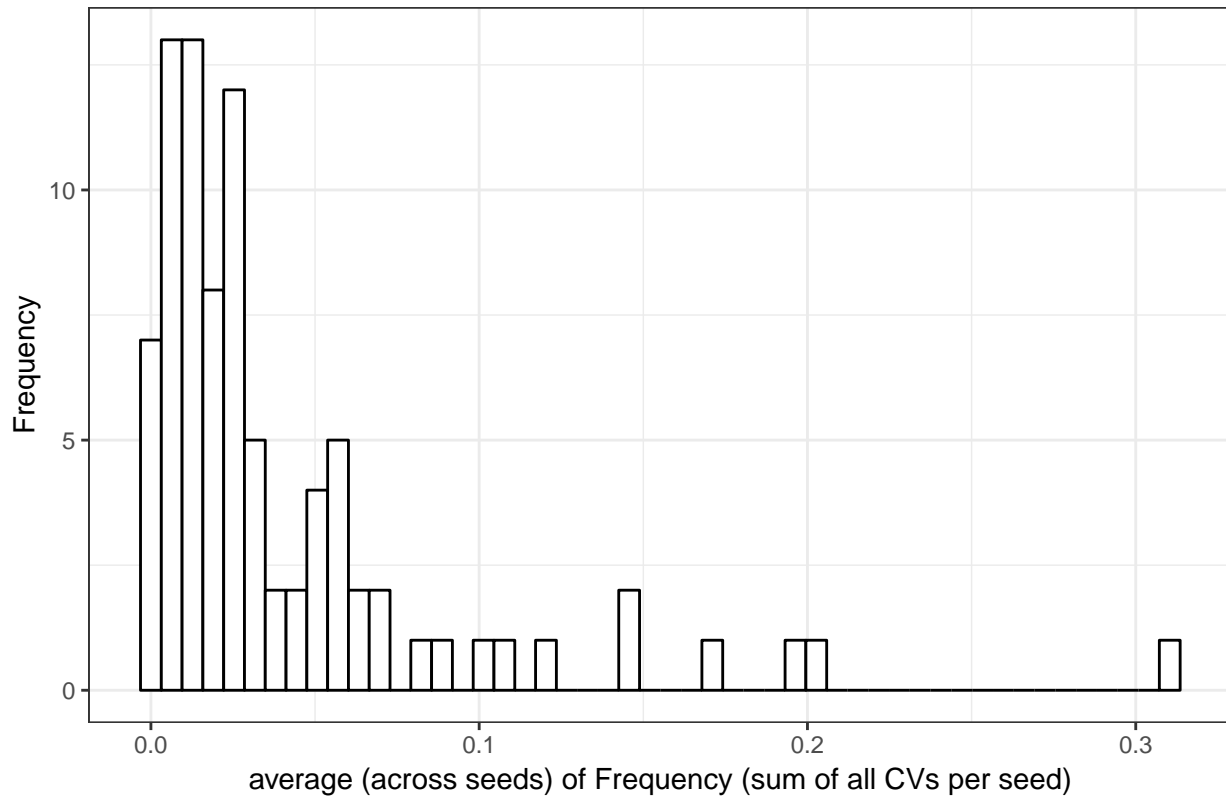
Averaged scores during RFE
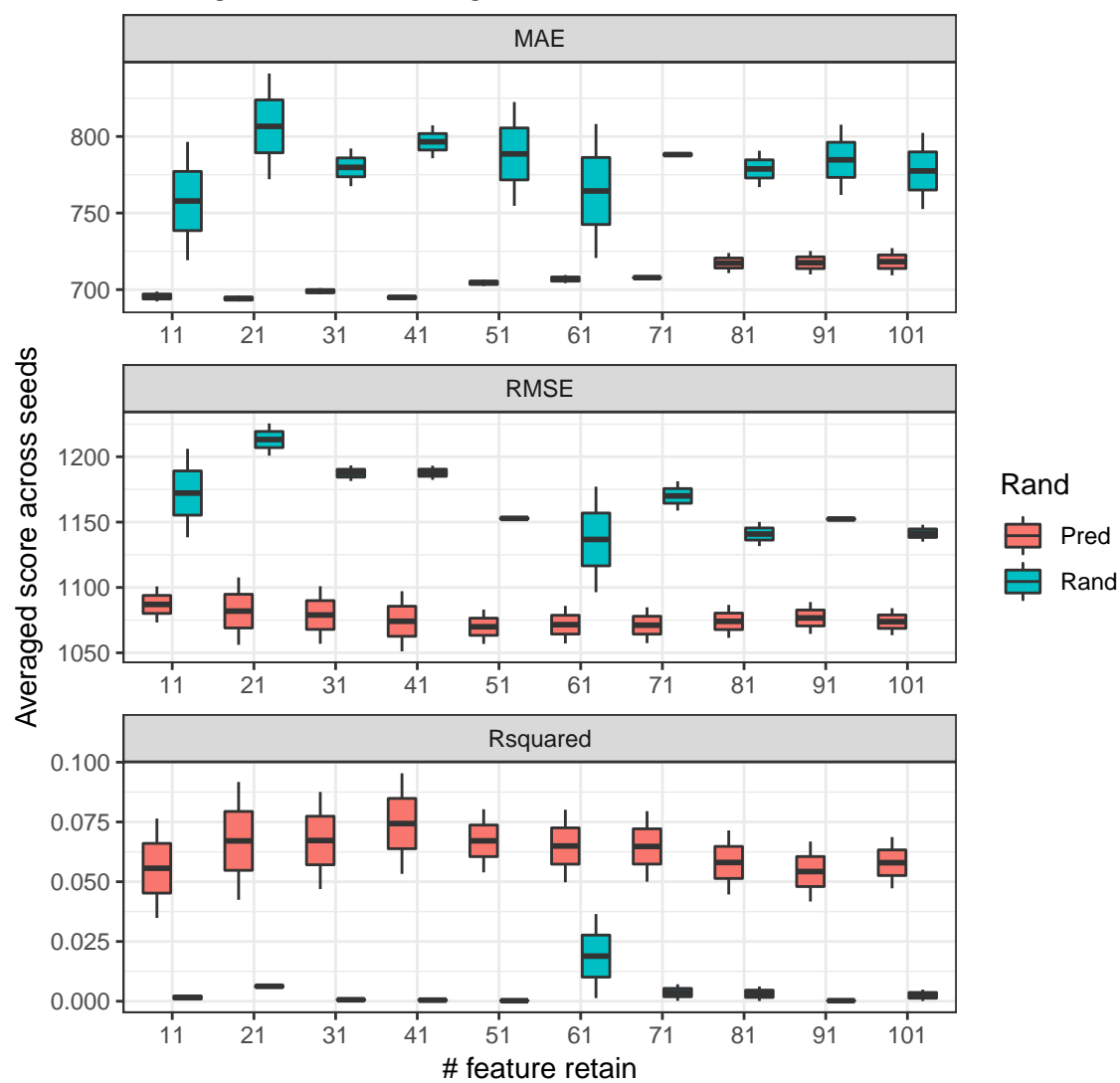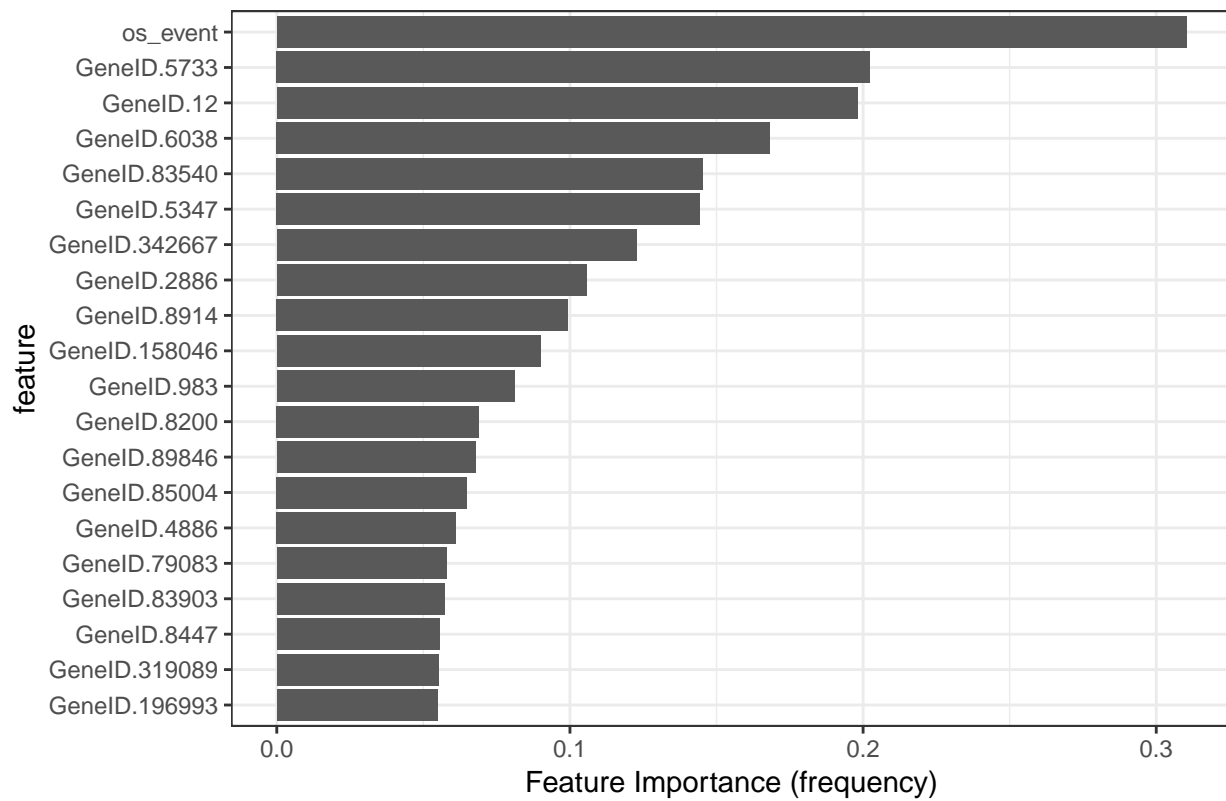
correlation


```
## # A tibble: 2 x 3
##    seed cor.avg cor.sdt
##   <dbl>   <dbl>   <dbl>
## 1  1001   0.215  0.0138
## 2  1002   0.282  0.0168
```
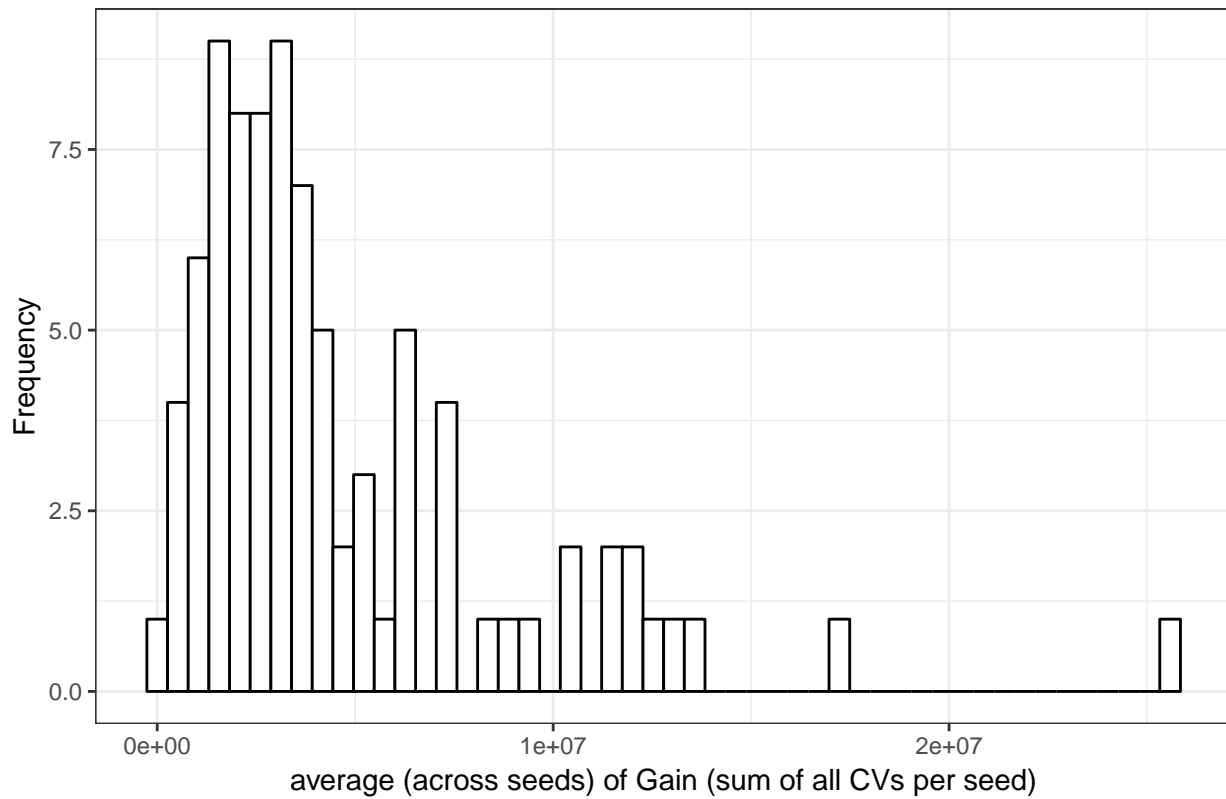
2. Important Features
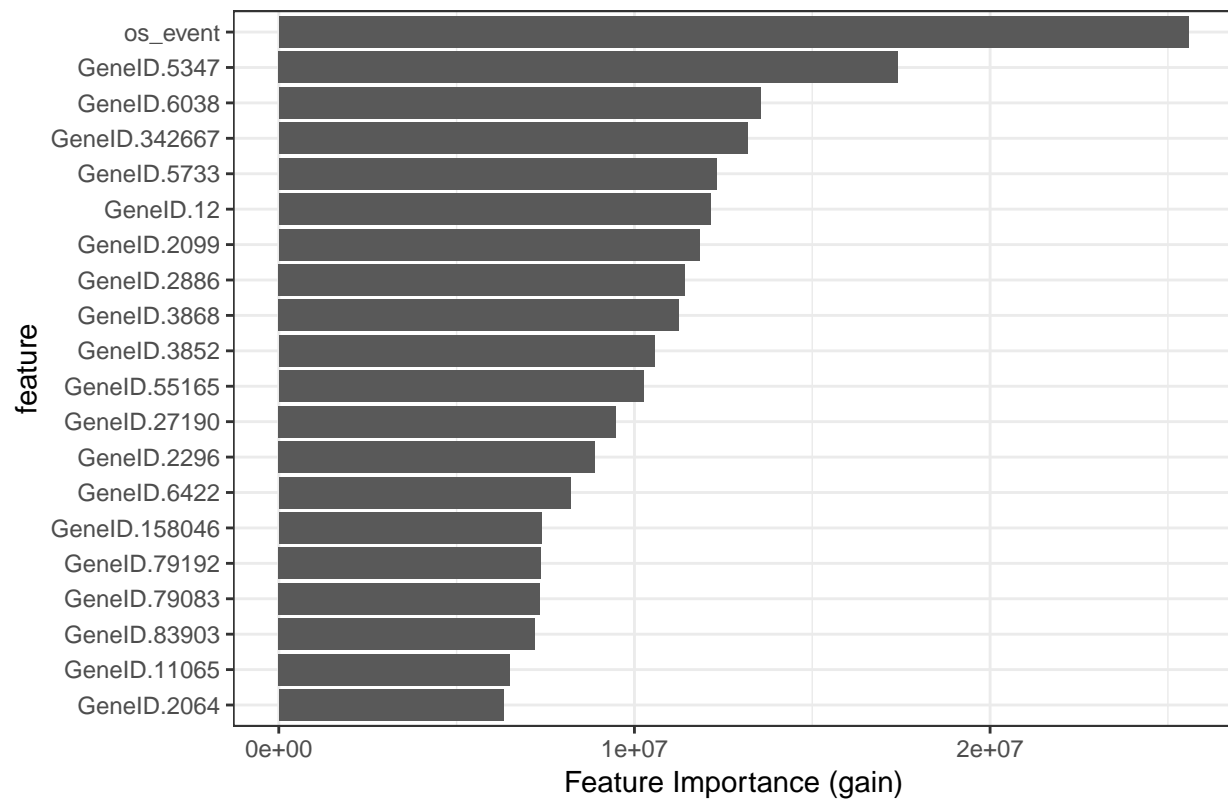
with 41 features based on Frequency

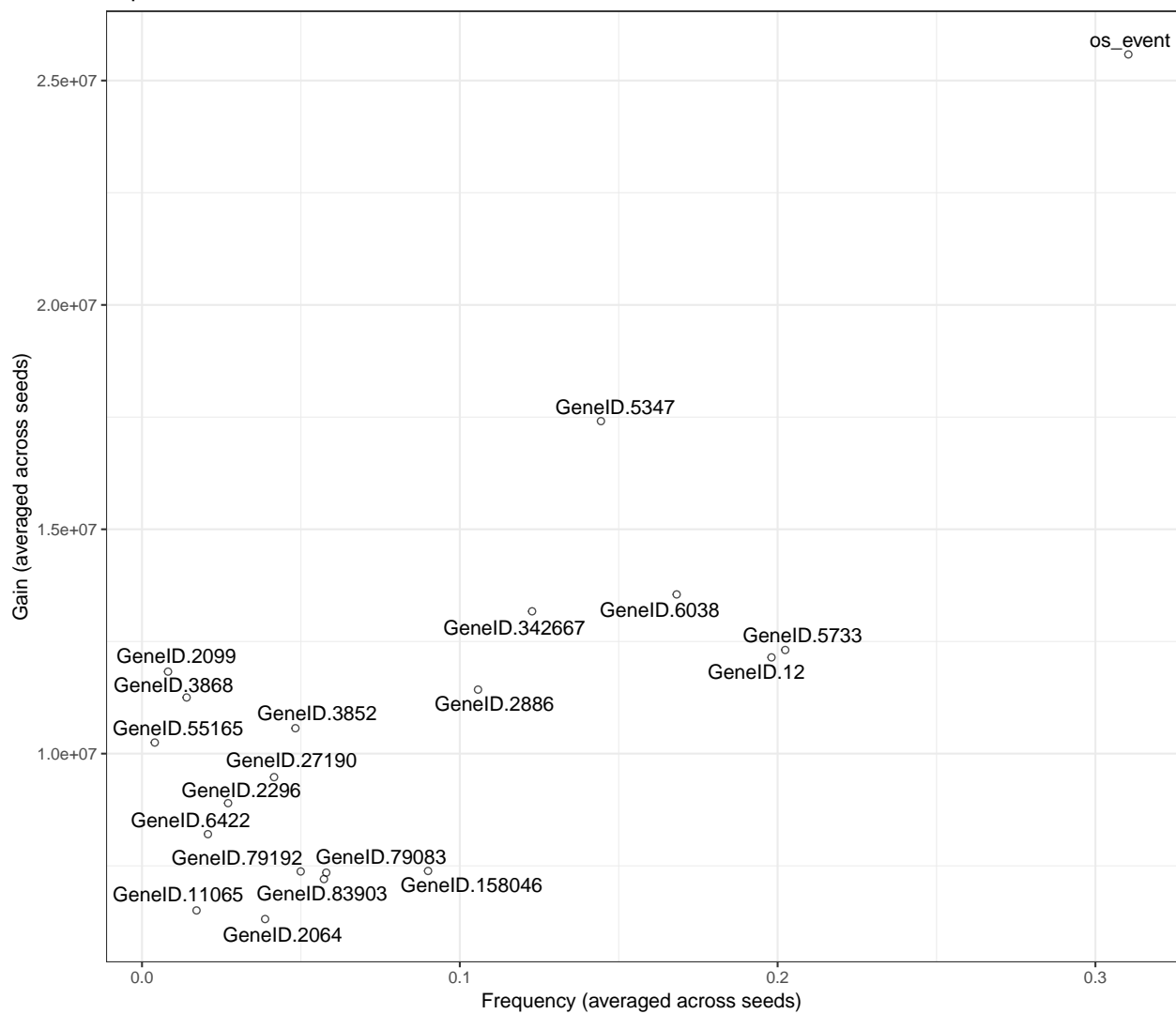## Top 20 features at 41 feature set based on Frequency

with 41 features based on Gain

Top 20 features at 41 feature set based on Gain

Top 20 features at 41 feature set

3. Hyper-parameters