Evaluate testing data (regression) - Lasso

EVE W.

2020-04-18
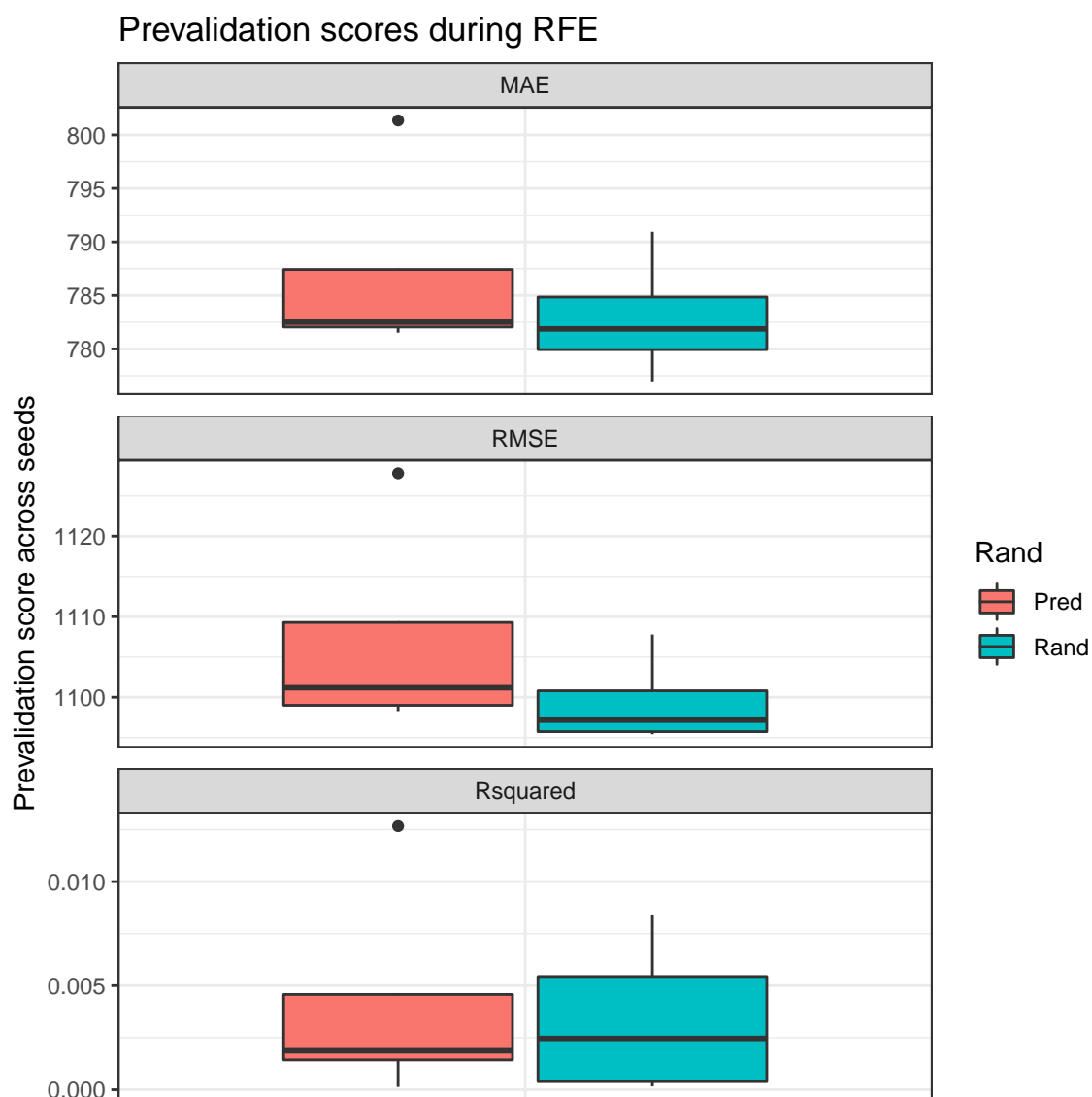
Contents

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

0. Load Data

```
## Error : $ operator is invalid for atomic vectors
```

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

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

```
## 4 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
```

run with lasso.r.

1. Scores
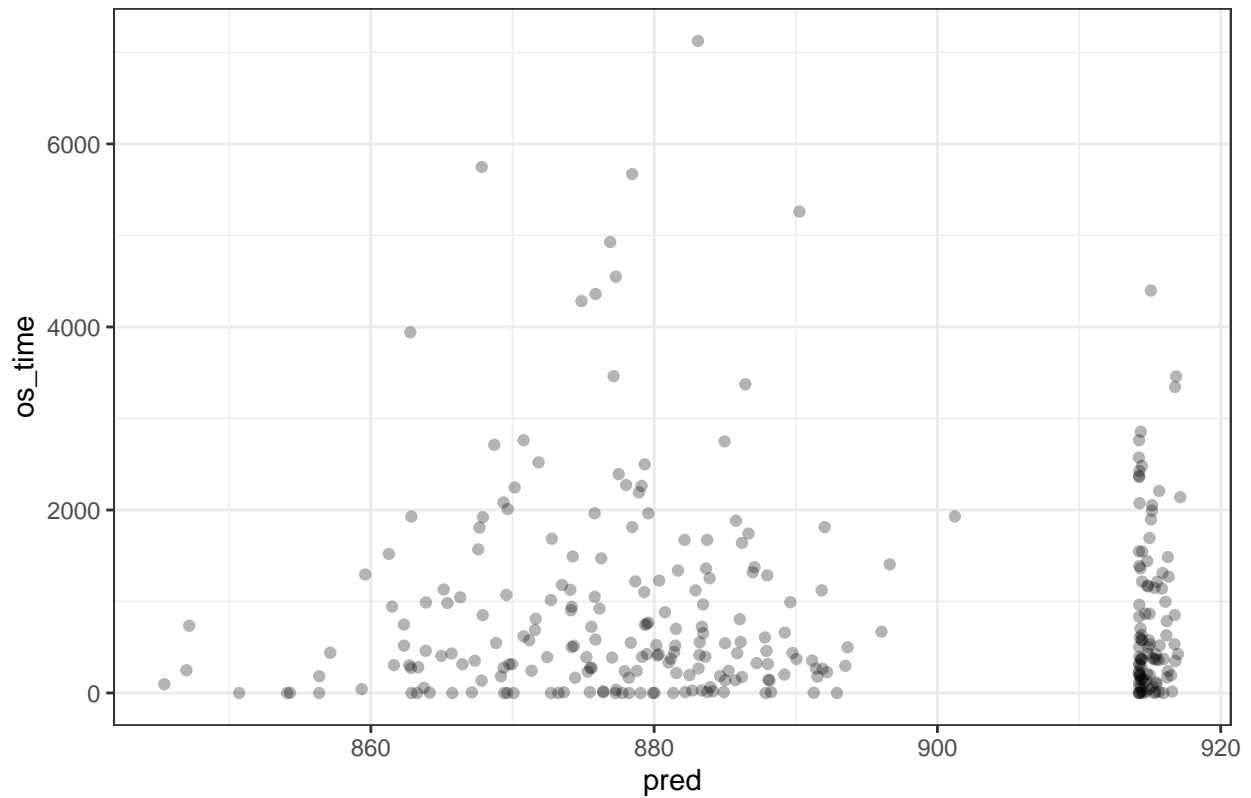
Prevalidation scores during RFE

'Pred' compares the actual CV prediction with observed value. 'Rand' compares permuted CV prediction with observed to mimic random prediction.
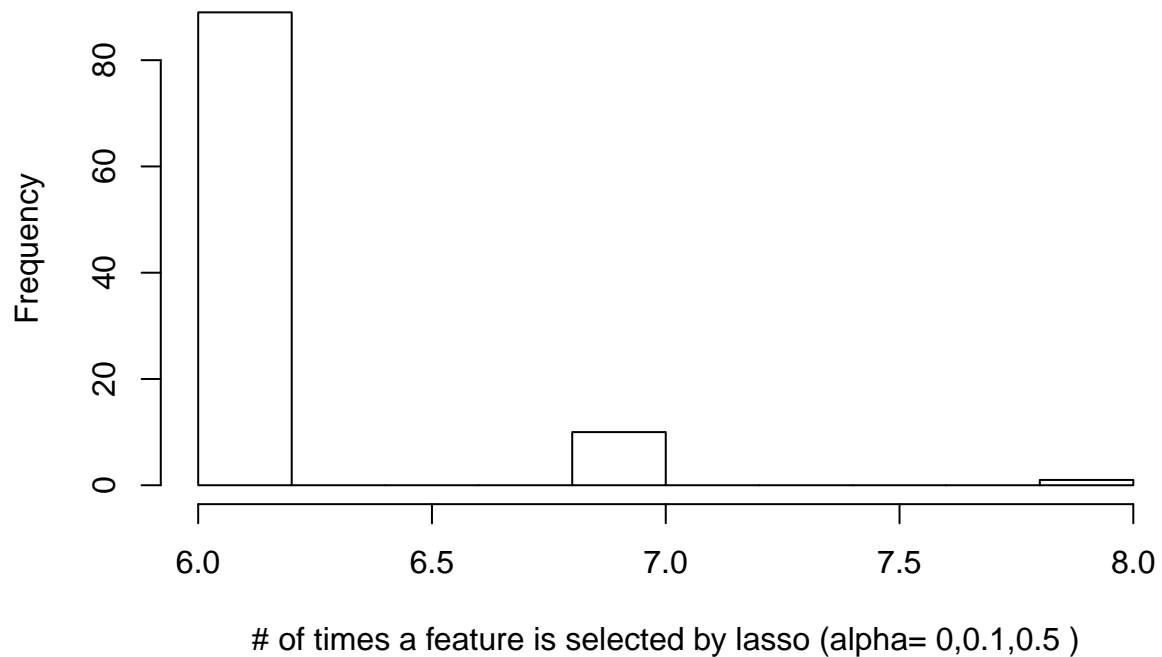
correlation

```
##
##
## Table: Averaged pearson correlation across seeds
##
##     cor.avg      cor.sdt
## -----------   ----------
##  -0.0526536    0.0426274
```

Correlation at seed = 1001 using 100 feature set input

2. Important Features

distribution across 3 seed x 3 CV

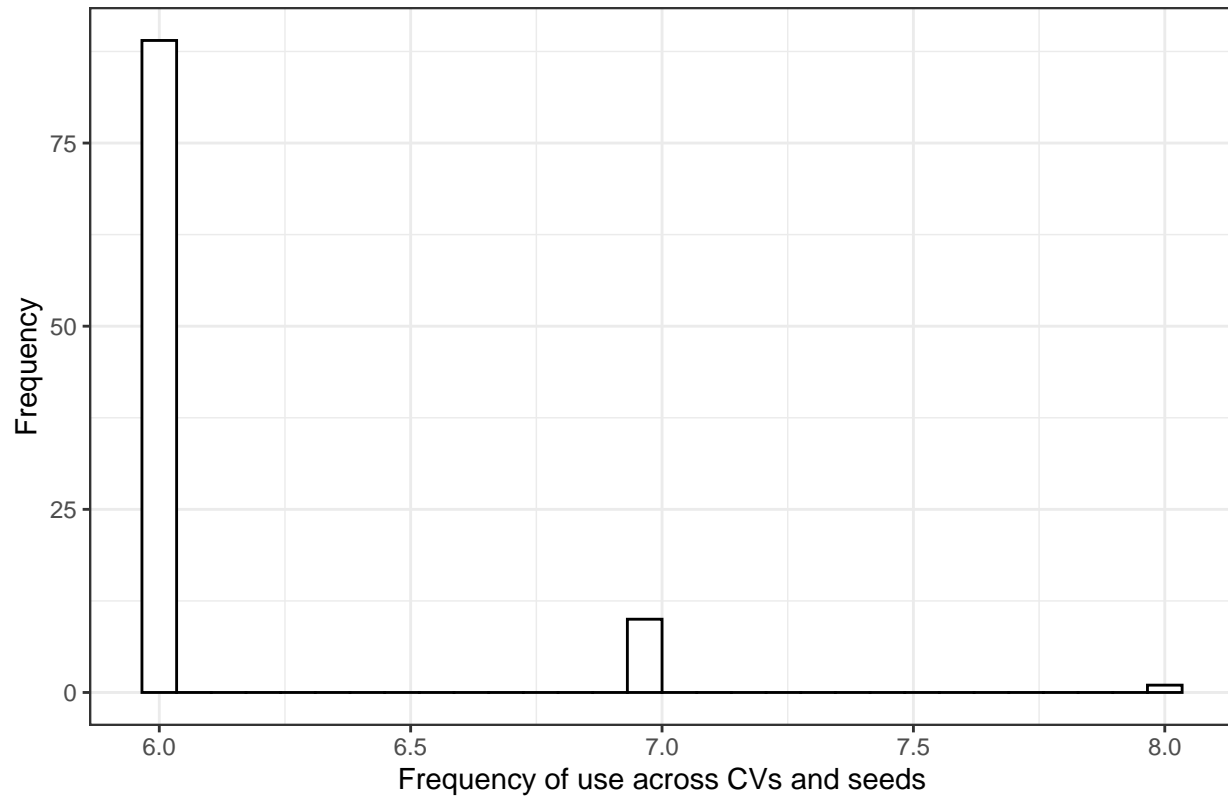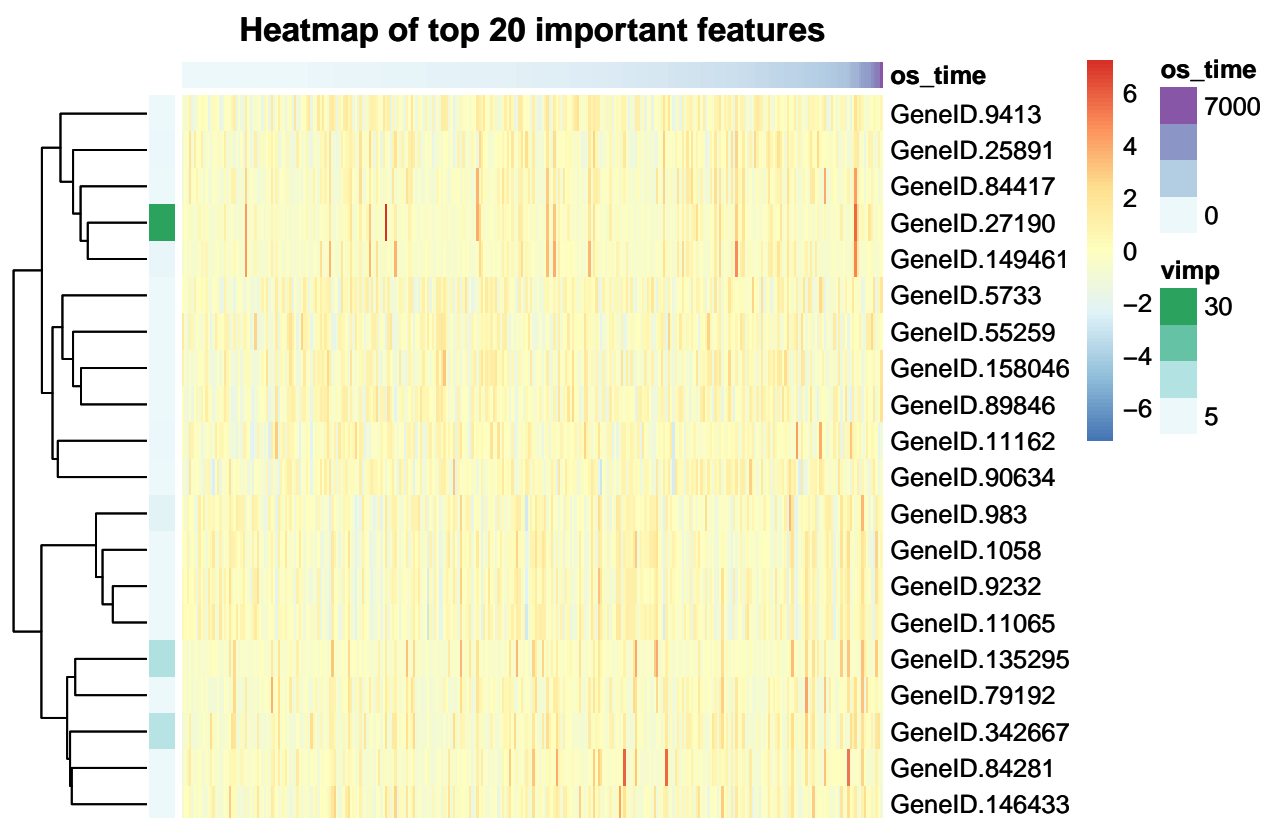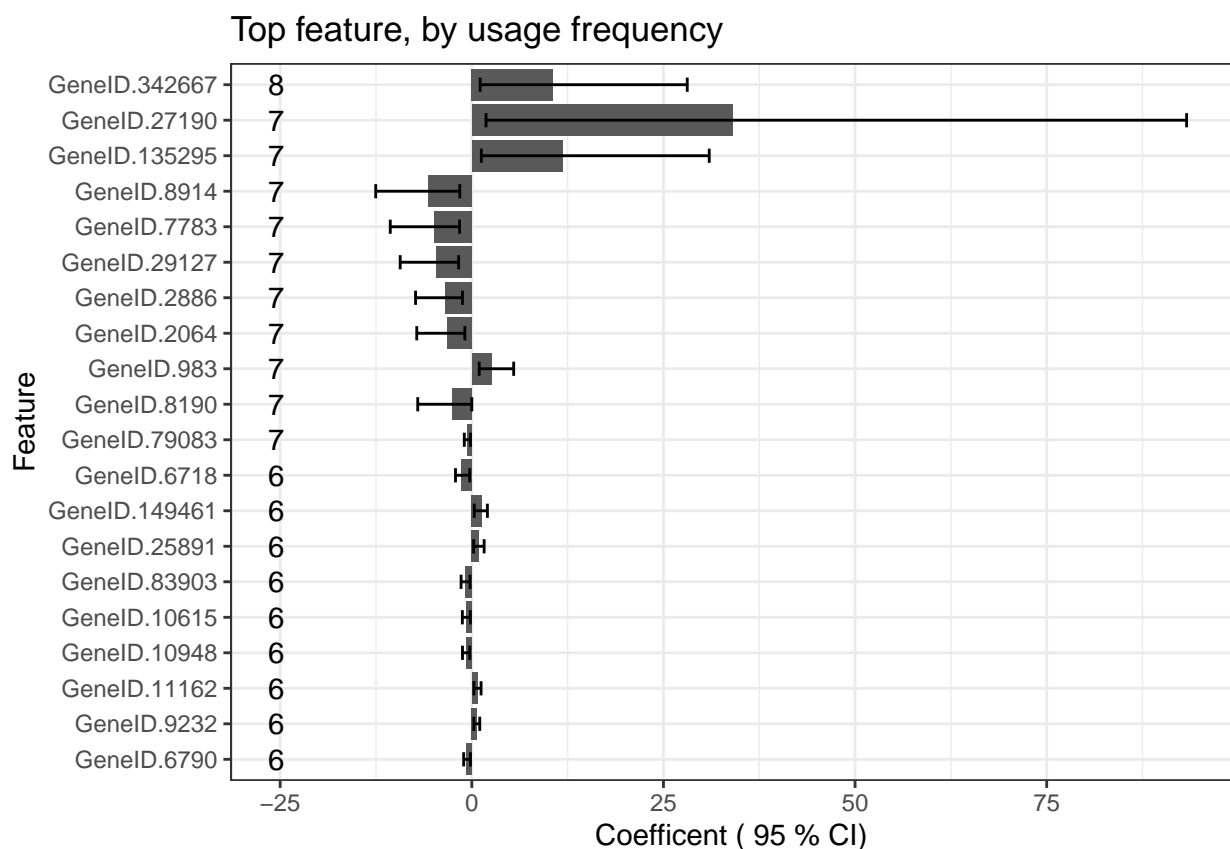# of times a feature is selected by lasso (alpha= 0,0.1,0.5)

```
## [1] "there are 100 unique features used from the 100 feature set"
## [1] "summary of number of features used in each run under 3 seeds and 3 CVs"

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       1      10     100      68     100     100

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

Distribution across all 100 features

Frequency of use across CVs and seeds

Top feature, by usage frequency

Heatmap of top 20 important features

Coefficients are labeled as vimp in the legend.