Evaluate testing data (regression) - Lasso

EVE W.

2019-11-16

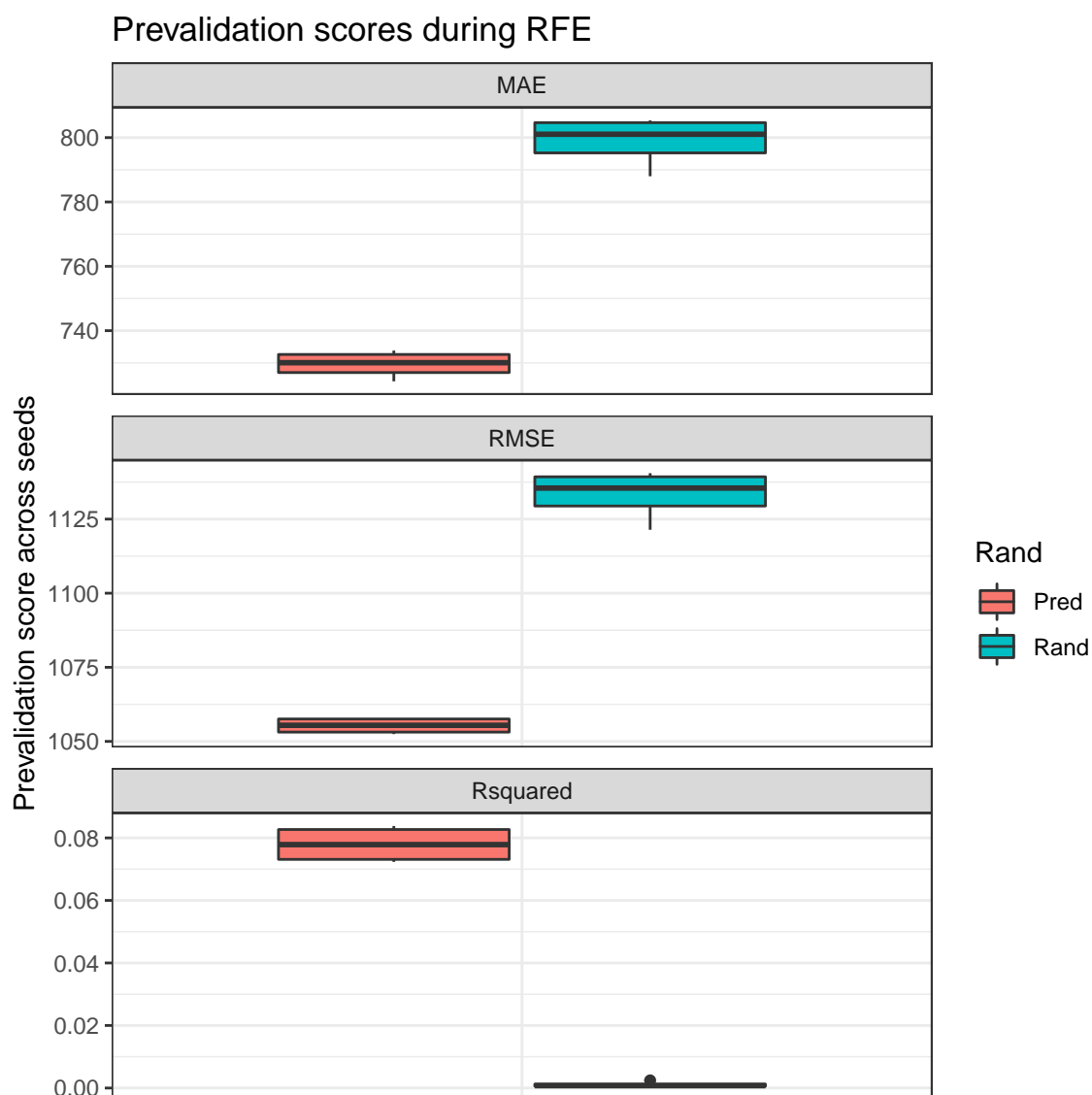## Contents

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

0. Load Data

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

```
## 101 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.
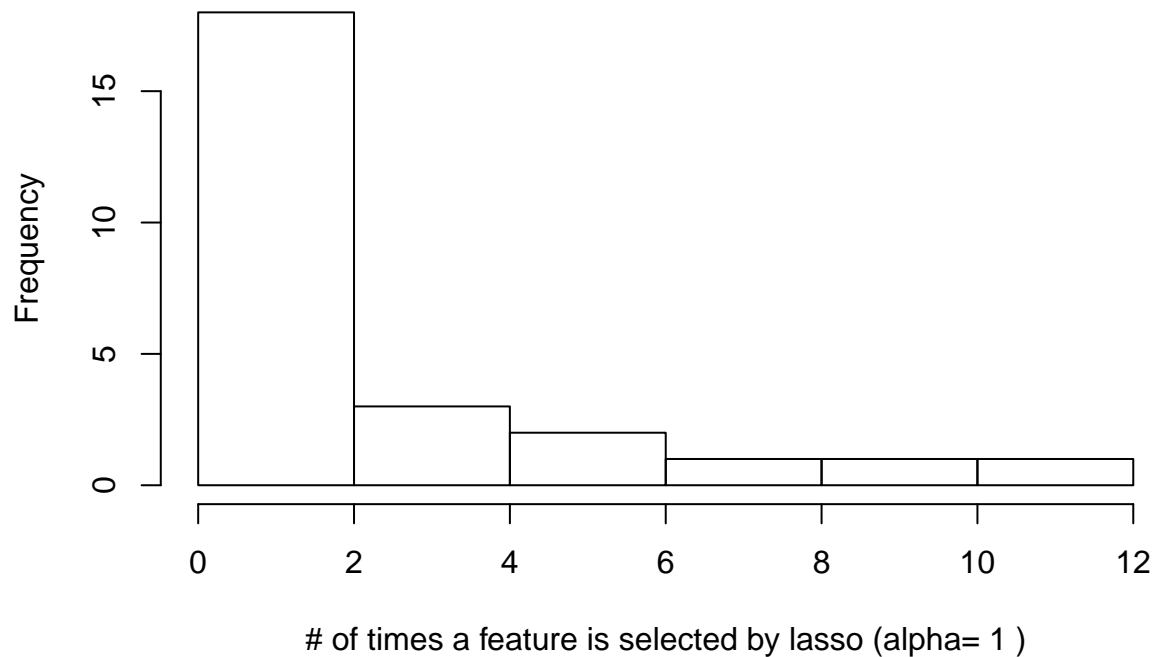
correlation

```
## 
## 
## Table: Averaged pearson correlation across seeds
## 
##    cor.avg      cor.sdt
## ----------   ----------
##  0.2790622    0.0106223
```
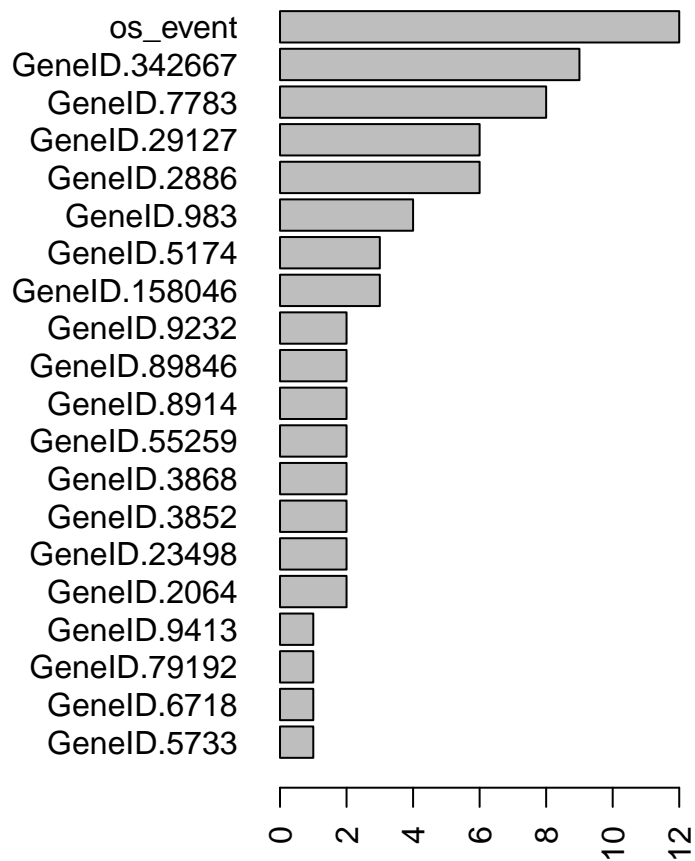
Correlation at seed = 1003 using 101 feature set input

os_time

6000

4000

2000

0

400   800   1200   1600

pred
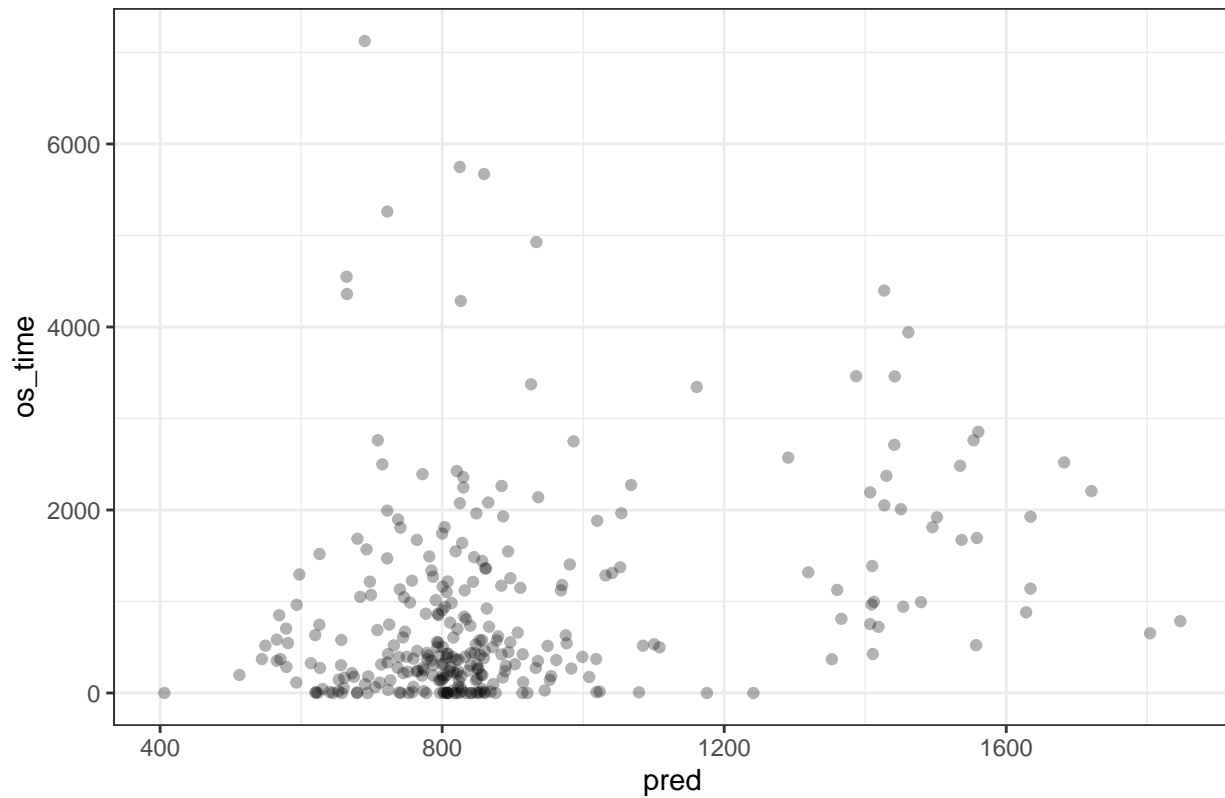
2. Important Features

distribution across 4 seed x 3 CV

Frequency

15

10

5

0

0   2   4   6   8   10   12

# of times a feature is selected by lasso (alpha= 1)

Number of times a feature is use


```
## (currently only Lasso has this graph)[1] "there are 26 unique features used from the 101 feature set
## [1] "summary of number of features used in each run under 4 seeds and 3 CVs"

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.000   4.000   6.000   6.417   7.250  14.000
```

Top Features

Heatmap of top 20 important features