Evaluate testing data (regression) - Lasso

John Doe

2019-01-30

Contents

Labels: os_time

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

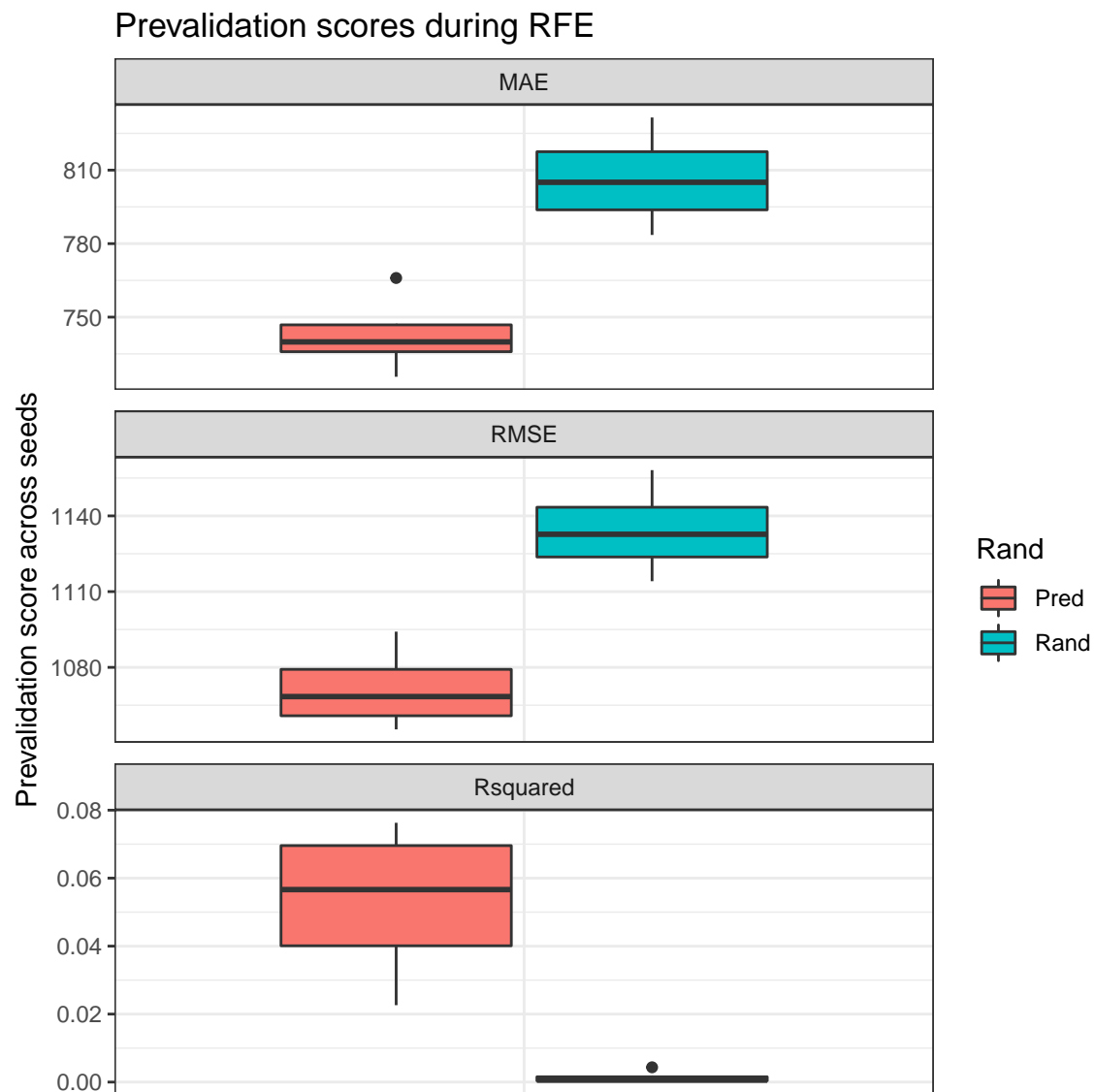## 0. Load Data

```
## 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
```

1. **Scores**

Prevalidation scores during RFE
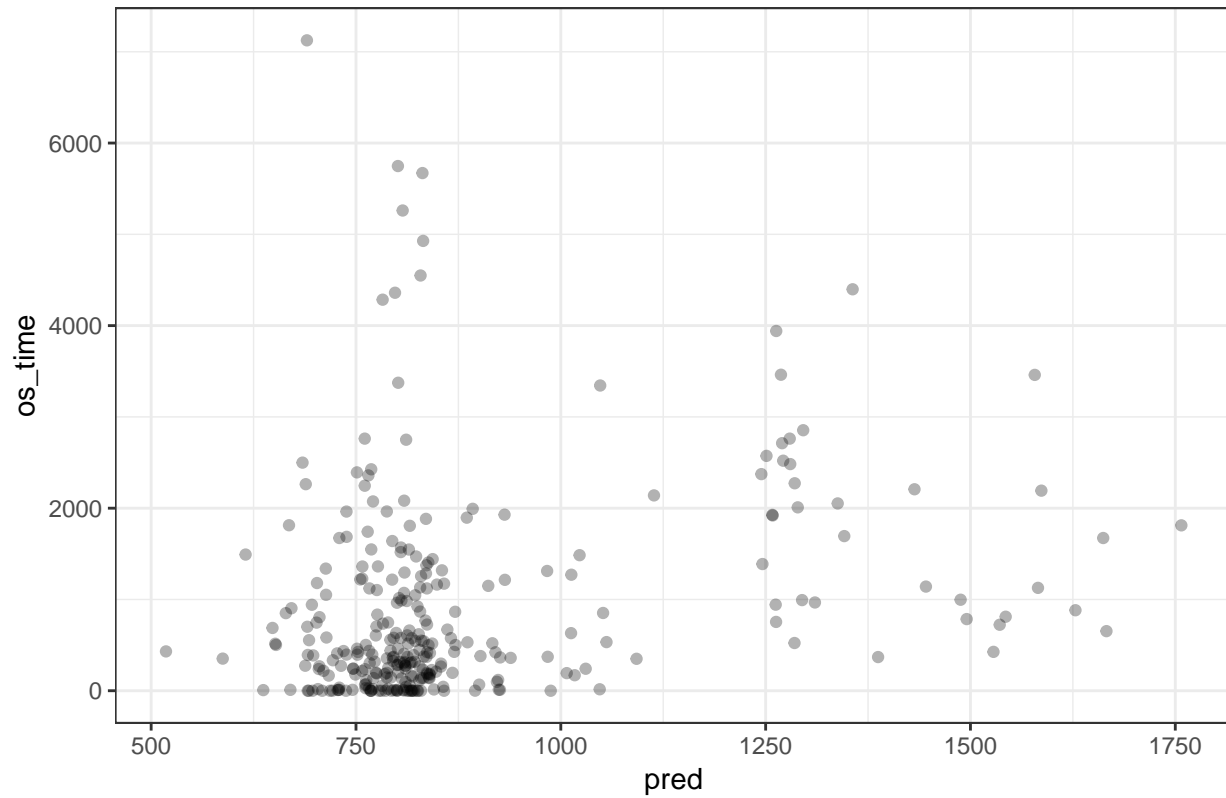
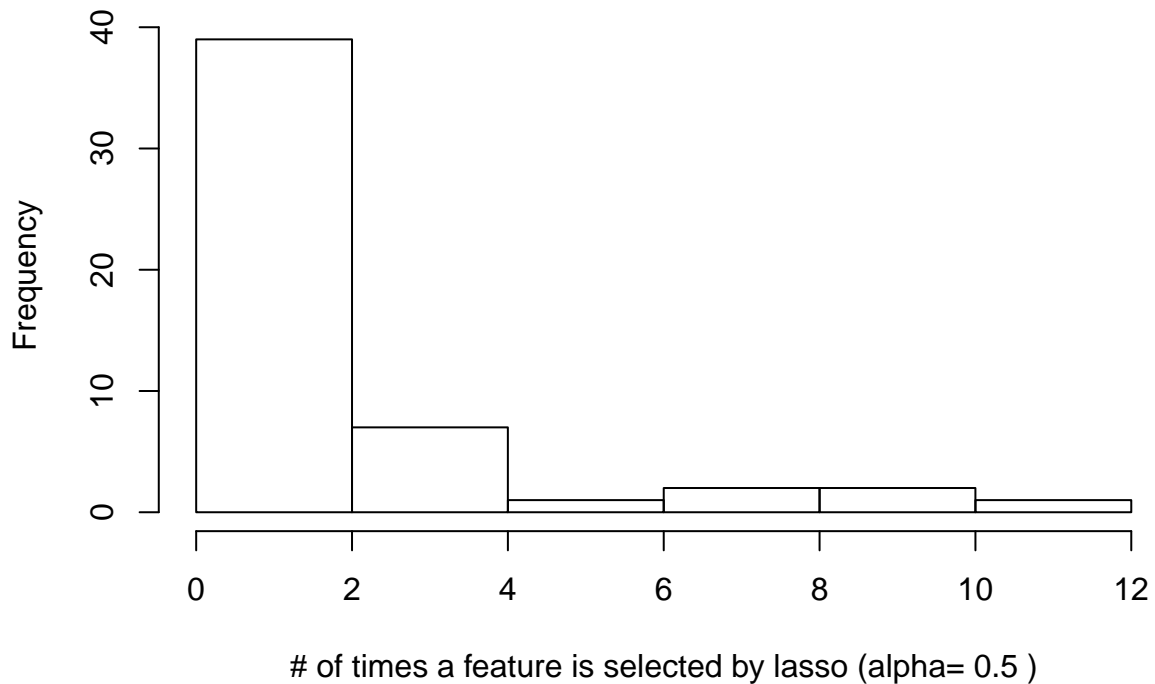

correlation

```
## Averaged pearson correlation across seeds# A tibble: 1 x 2
##   cor.avg cor.sdt
##     <dbl>   <dbl>
## 1   0.225  0.0562
```
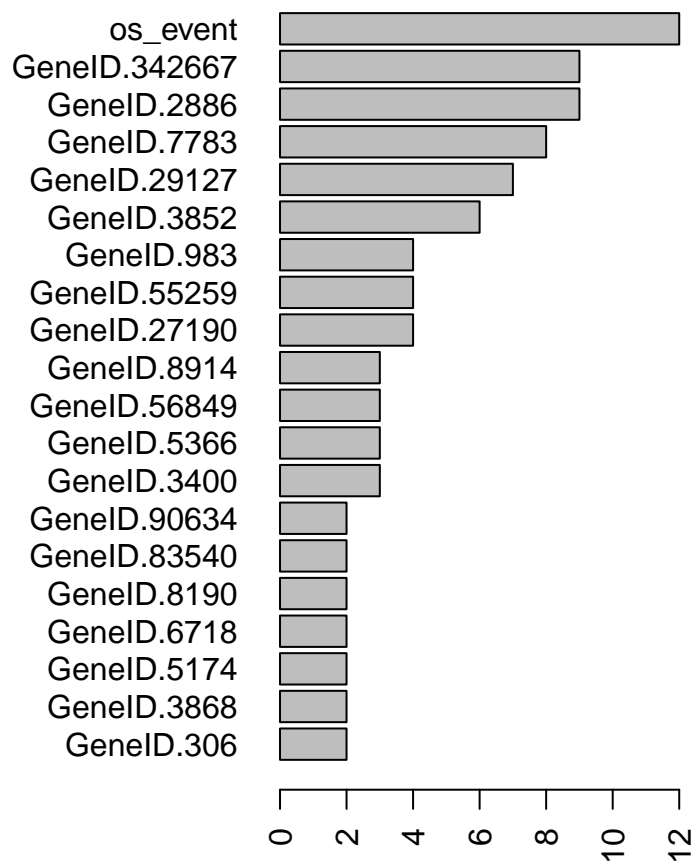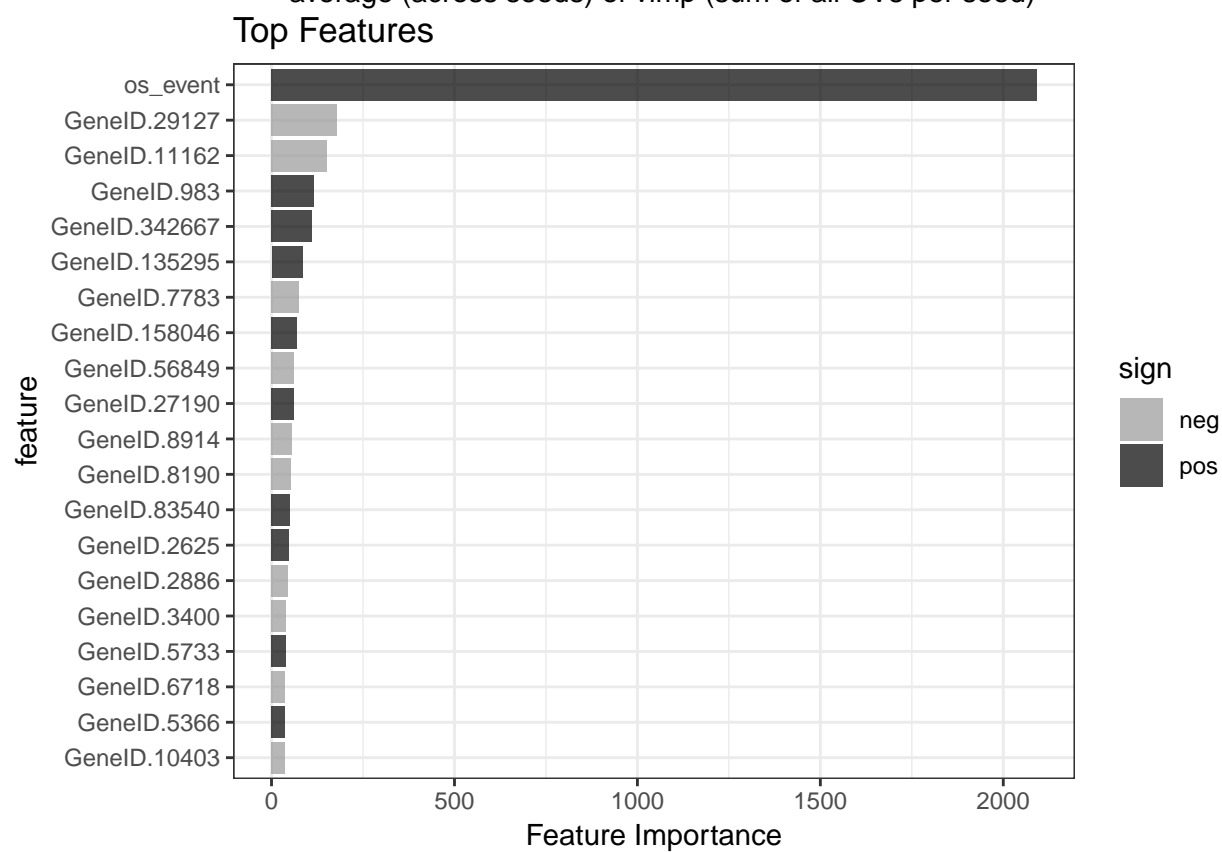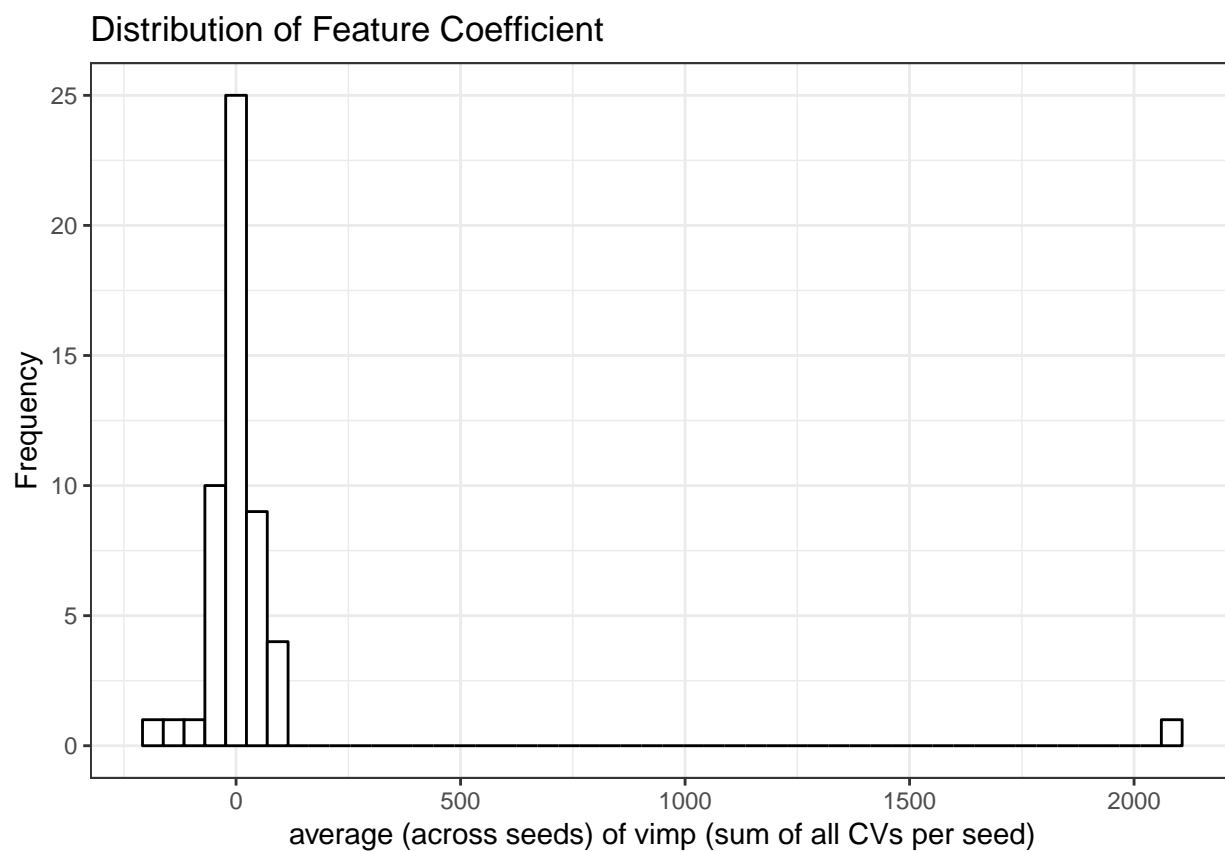
Correlation at seed = 1001

2. Important Features

distribution across 4 seed x 3 CV

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

Number of times a feature is used

os_event
GeneID.342667
GeneID.2886
GeneID.7783
GeneID.29127
GeneID.3852
GeneID.983
GeneID.55259
GeneID.27190
GeneID.8914
GeneID.56849
GeneID.5366
GeneID.3400
GeneID.90634
GeneID.83540
GeneID.8190
GeneID.6718
GeneID.5174
GeneID.3868
GeneID.306

0 2 4 6 8 10 12

(currently only Lasso has this graph)

Distribution of Feature Coefficient

Top Features

Heatmap of top 20 important features