Evaluate testing data (regression) - Lasso

EVE W.

2020-04-20

Contents

0. Load Data 1
1. Scores 2
correlation 2
2. Important Features 3

```
## user input
project_home <- "~/EVE/examples"
project_name <- "lasso_regression2"
```

0. Load Data

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

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

```
## 101 of full features
```

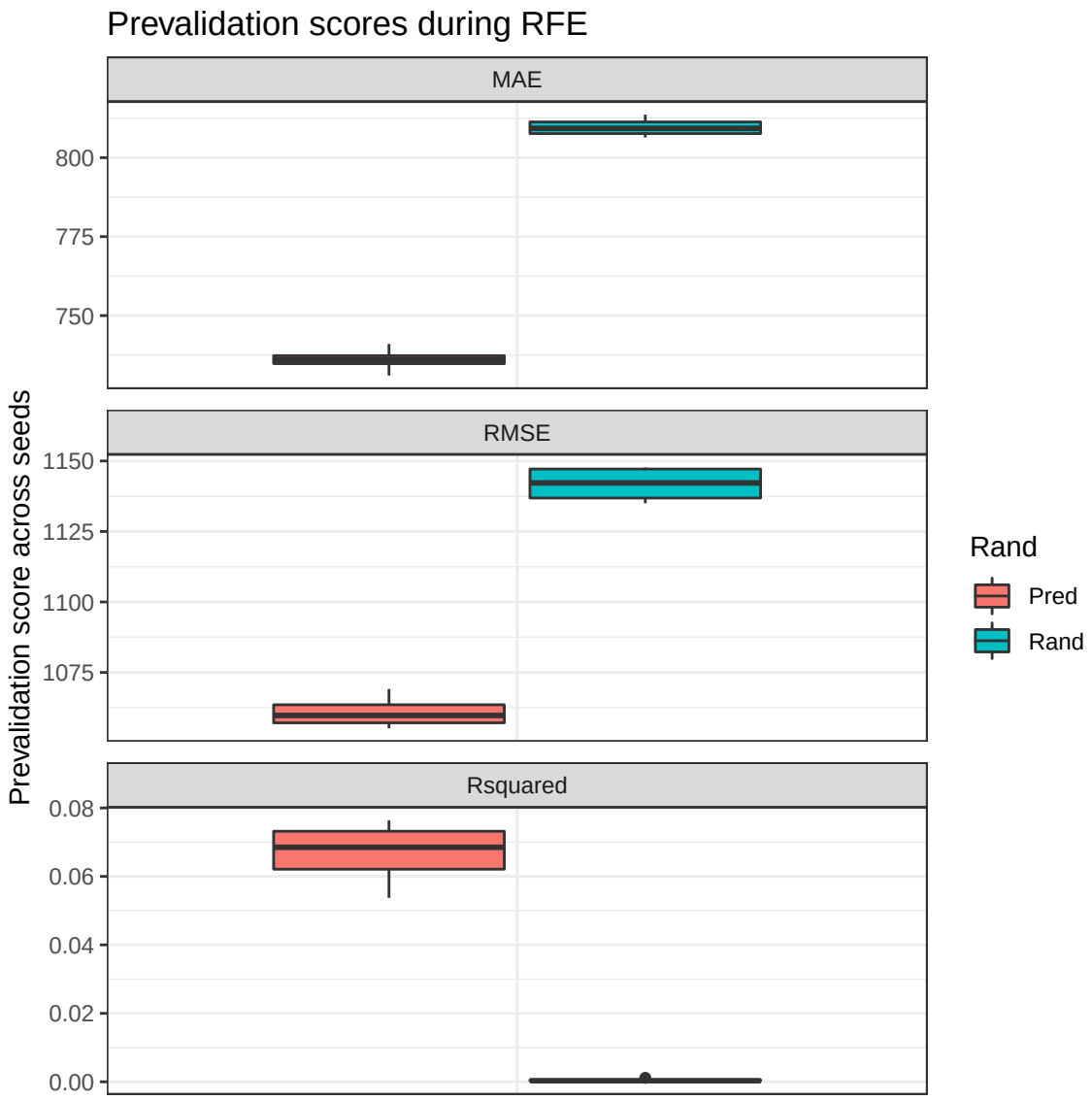
```
## 4 runs, each run contains 5 CVs.
```

```
## 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 with alpha = 0, 0.5, 1.

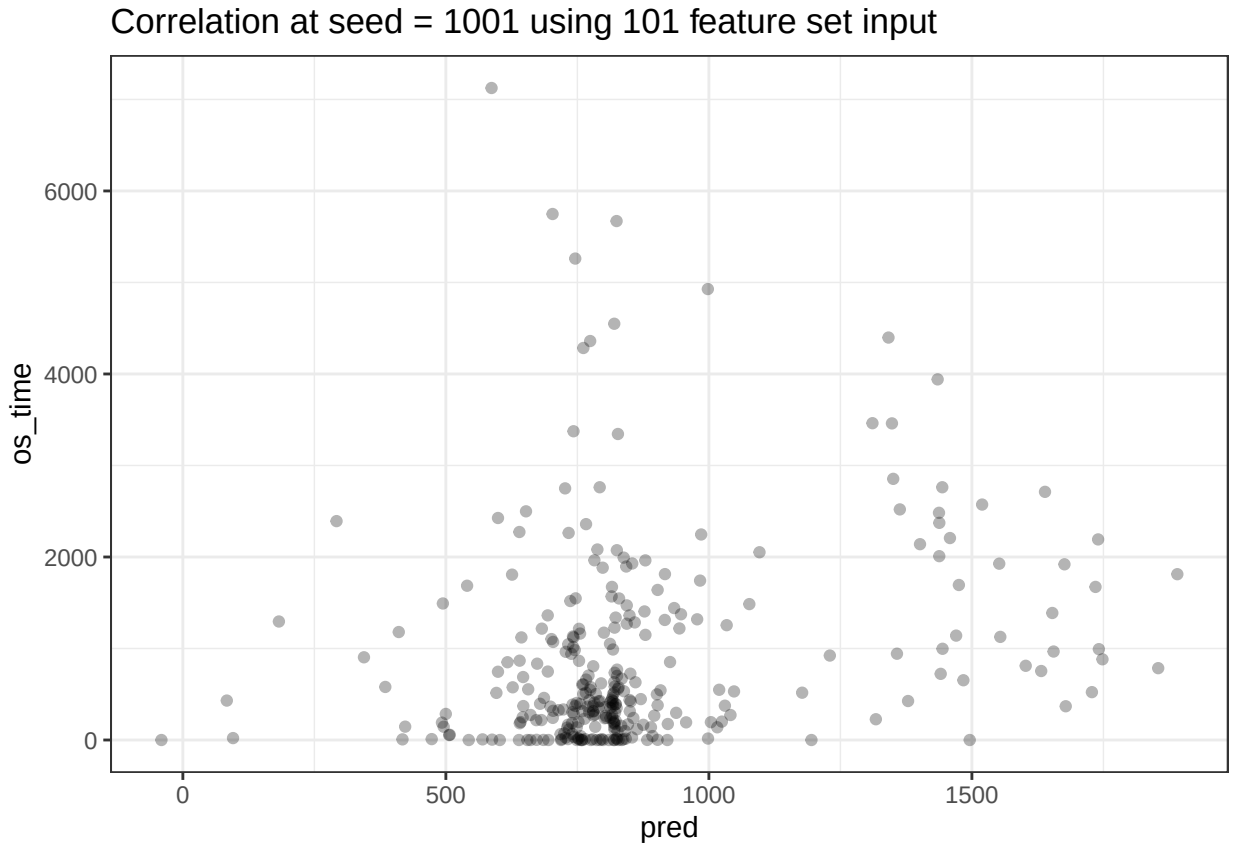
1. Scores



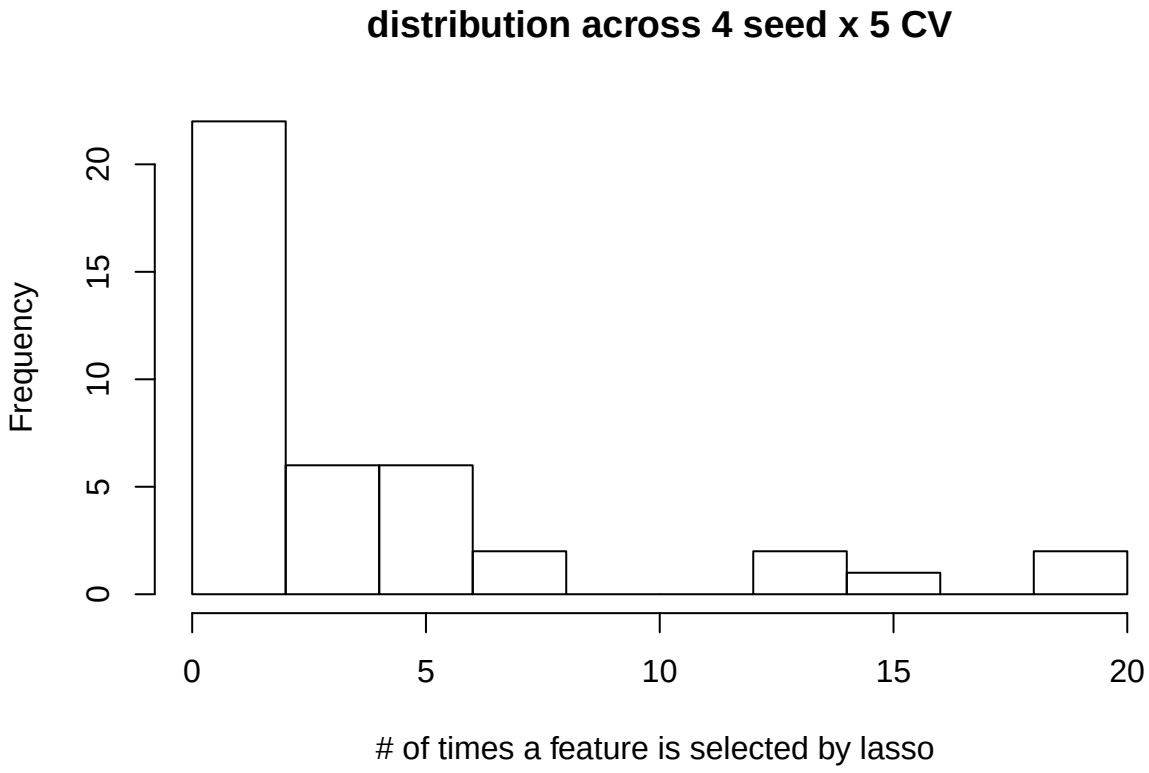
'Pred' compares the actual CV prediction with observed value. 'Rand' compares permuted CV prediction with observed to mimic random prediction.

correlation

```
##
##
## Table: Averaged pearson correlation across seeds
##
##    cor.avg     cor.sdt
## ----------  ----------
##  0.2578883   0.0195345
```



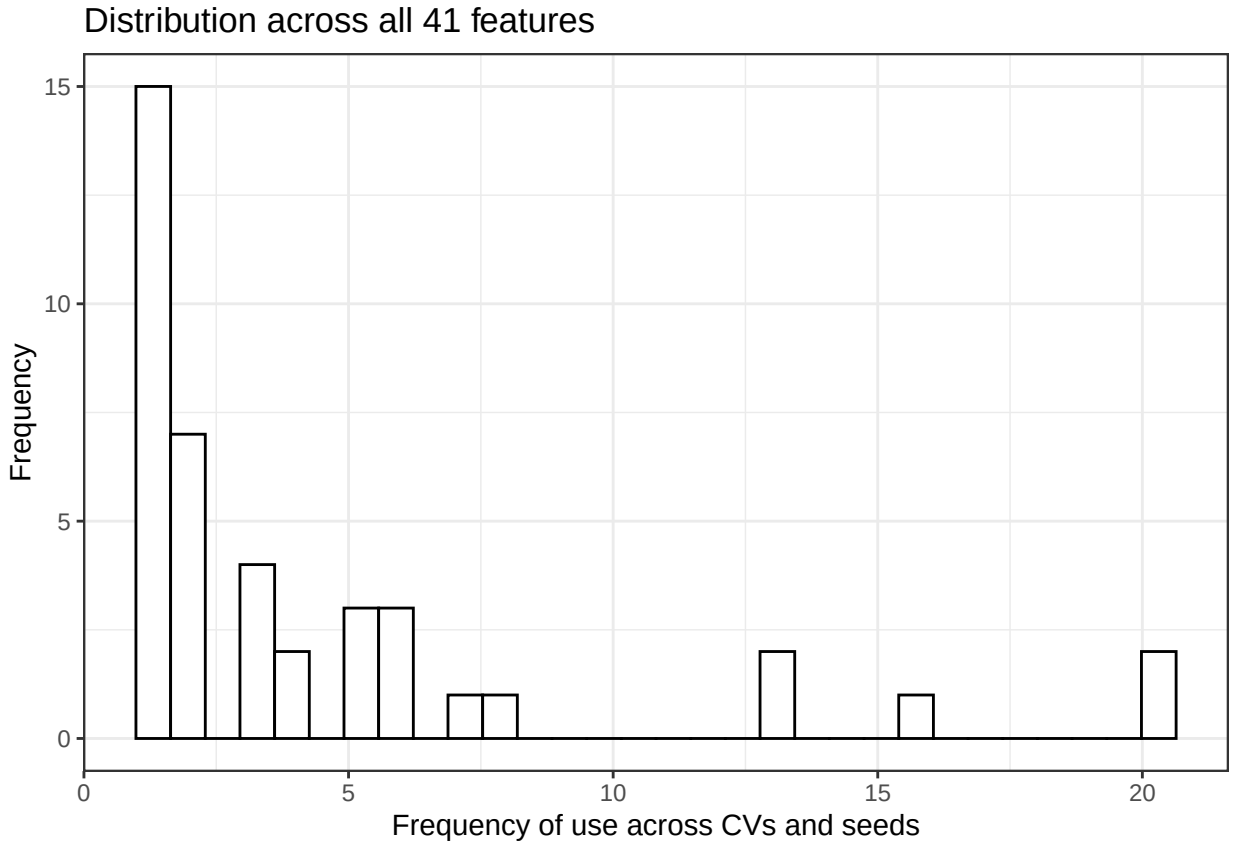
2. Important Features

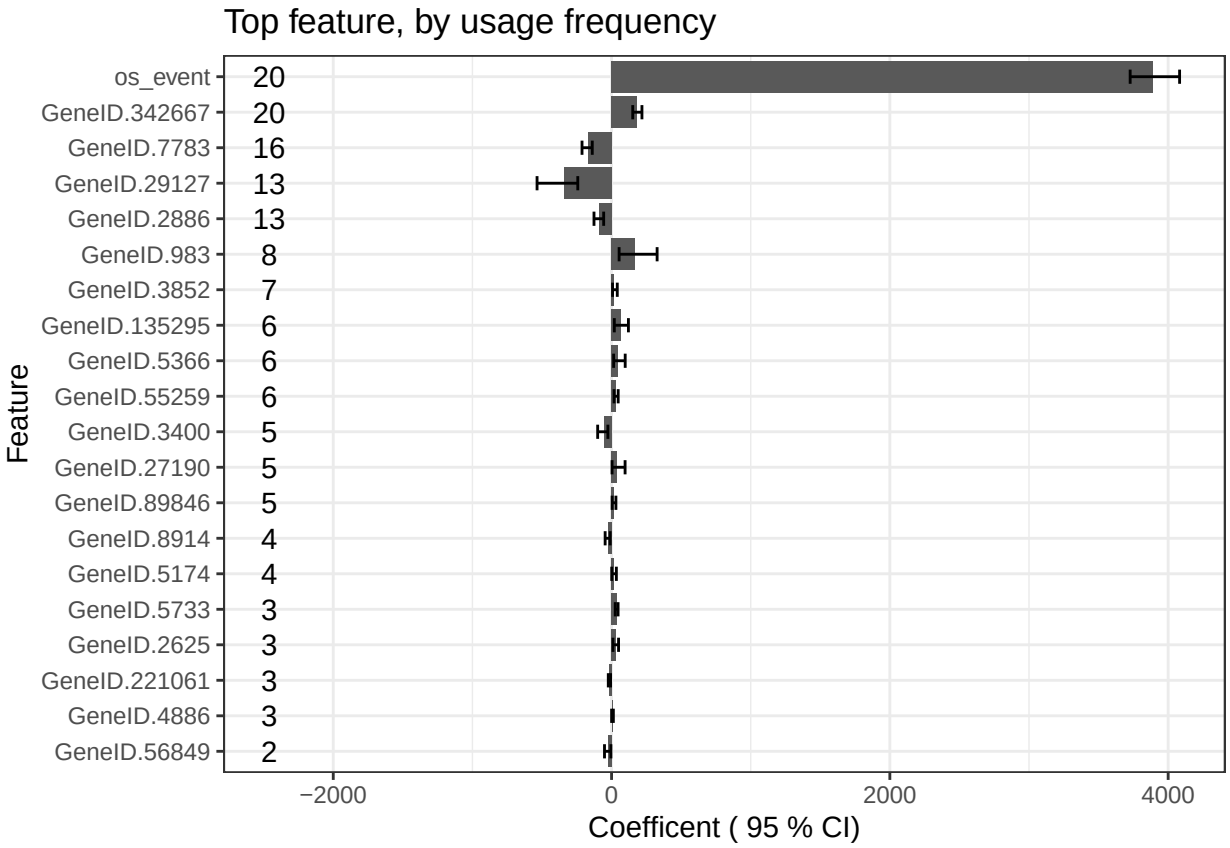


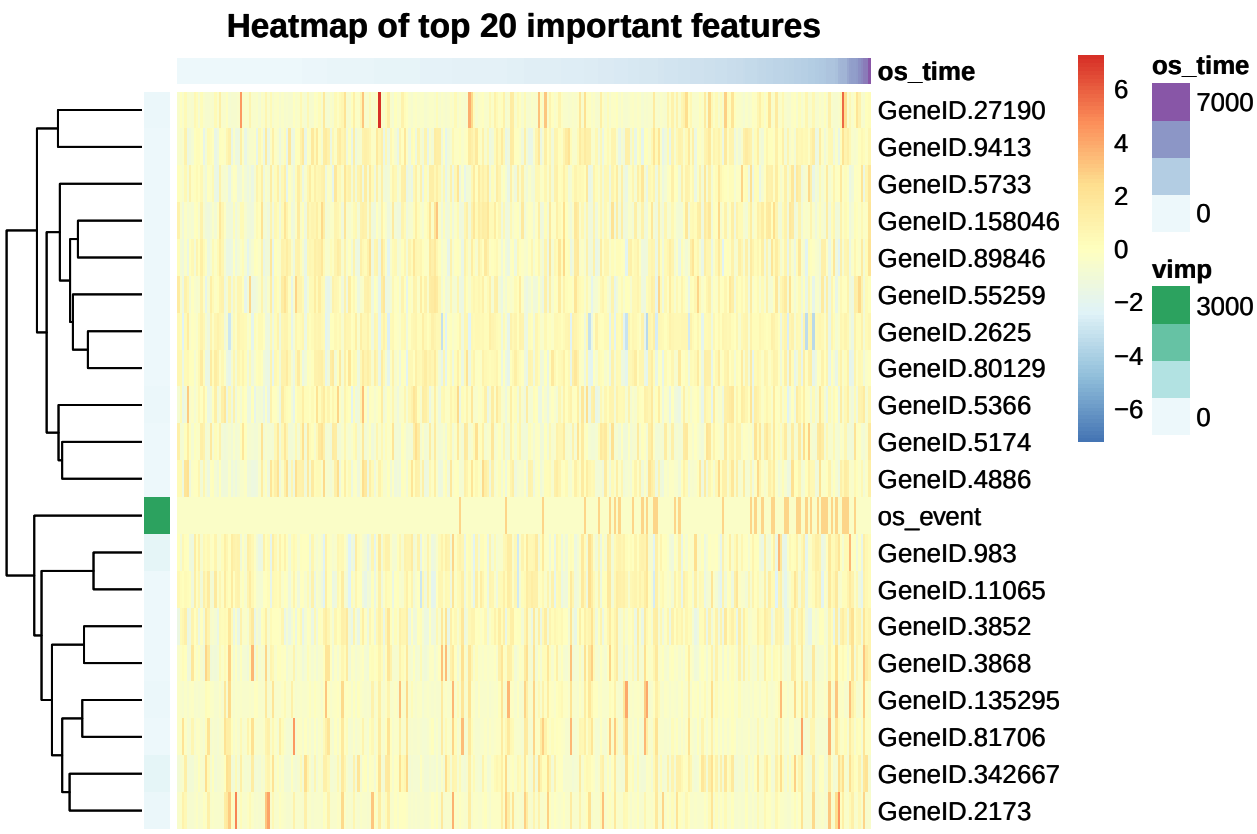
```
## [1] "there are 41 unique features used from the 101 feature set"
## [1] "summary of number of features used in each run under 4 seeds and 5 CVs"
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    3.00    4.75    8.00    8.95   10.75   28.00
```

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







lasso coefficients are labeled as 'vimp' above.

Table 1: parameter selection

seed	alpha	lambda	cv
1001	1.0	134.97480	1
1001	1.0	76.38207	2
1001	1.0	98.67150	3
1001	0.5	101.66514	4
1001	1.0	108.77792	5
1002	1.0	103.02430	1
1002	1.0	137.47989	2
1002	1.0	63.43829	3
1002	1.0	115.87089	4
1002	1.0	111.98319	5
1003	1.0	124.24167	1
1003	1.0	58.32844	2
1003	0.5	142.96597	3
1003	1.0	63.93014	4
1003	1.0	86.88993	5
1004	1.0	129.11759	1
1004	1.0	88.55919	2
1004	1.0	102.25761	3
1004	1.0	91.47572	4
1004	1.0	65.17750	5