

Evaluate testing data (regression) - Lasso

EVE W.

2019-07-30

Contents

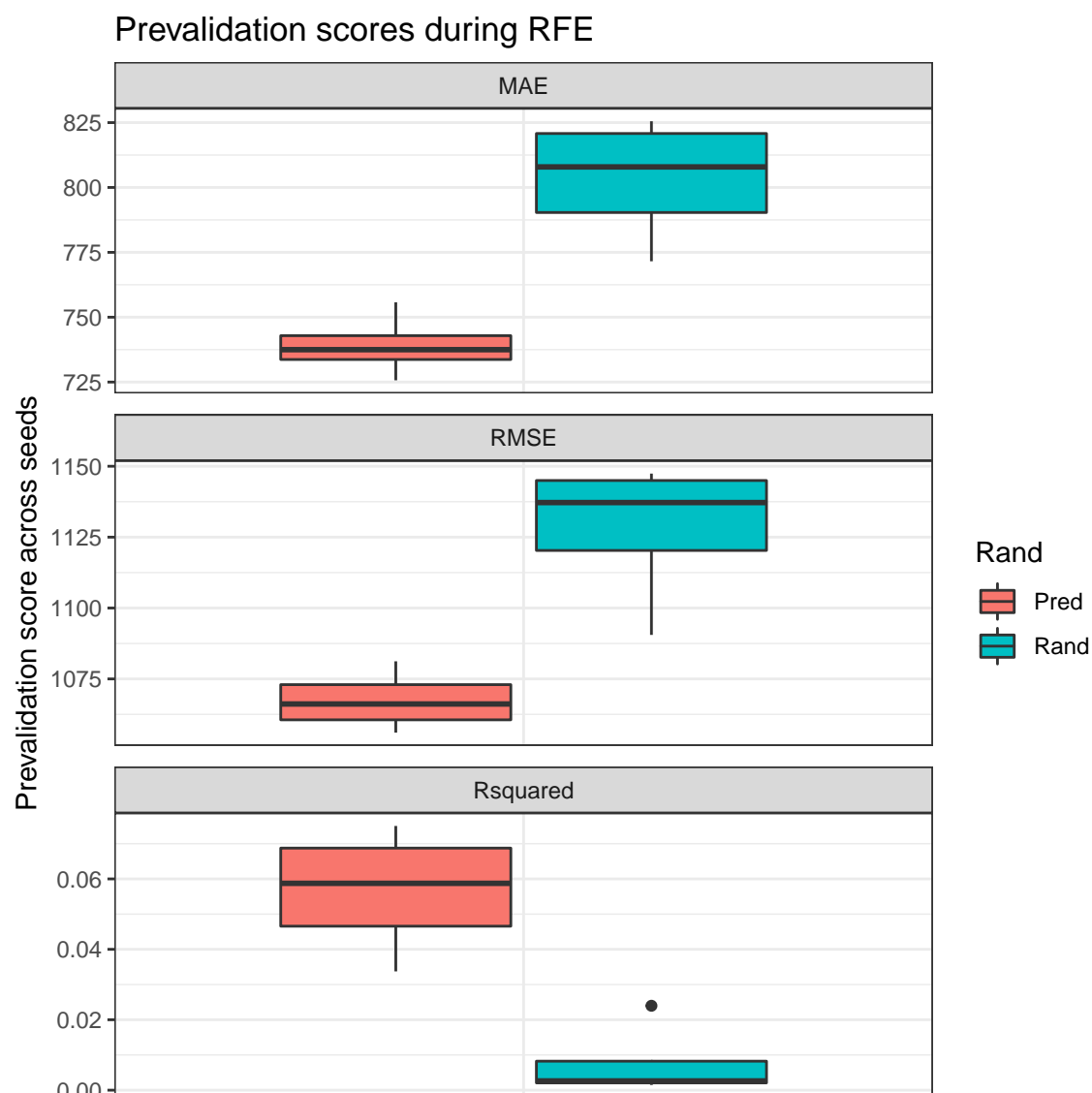
0. Load Data	1
1. Scores	2
correlation	2
2. Important Features	3

```
## 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

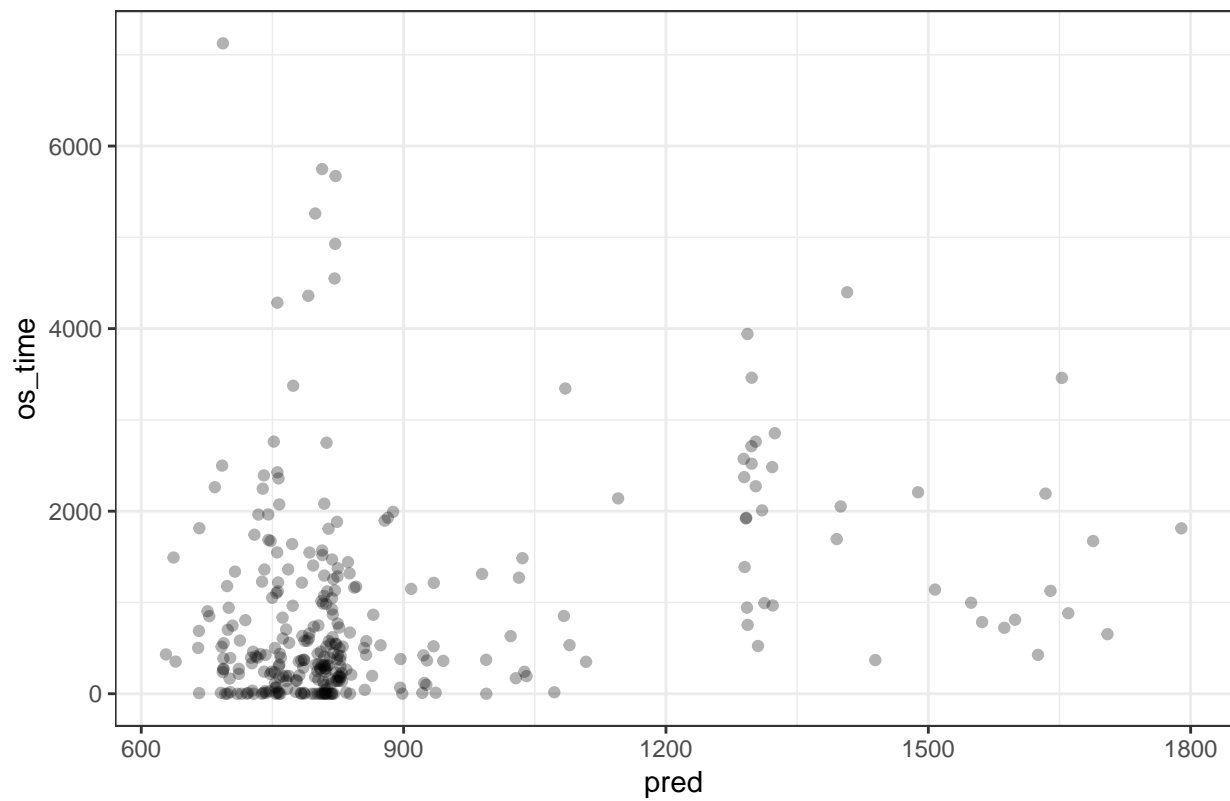


'Pred' compares the actual CV prediction with observed value. 'Rand' compares permuted CV prediction with observed to mimic random prediction.

correlation

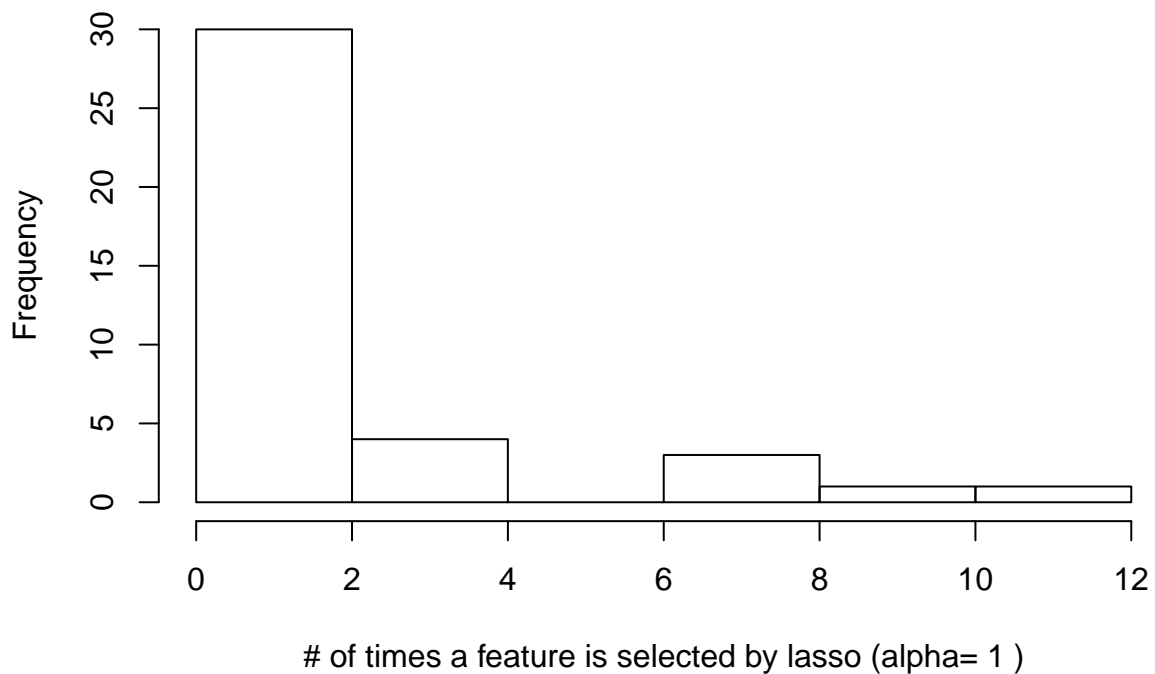
```
##
##
## Table: Averaged pearson correlation across seeds
##
##   cor.avg   cor.sdt
## -----
## 0.235311  0.0399068
```

Correlation at seed = 1001 using 101 feature set input

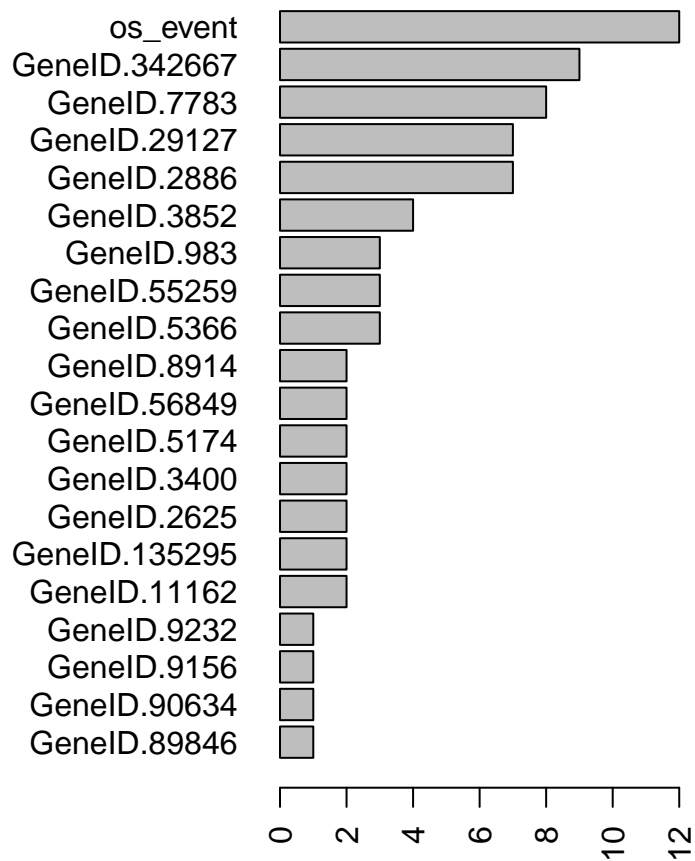


2. Important Features

distribution across 4 seed x 3 CV



Number of times a feature is used



(currently only Lasso has this graph)[1] "there are 39 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.
##	1.00	5.00	5.50	7.75	8.75	18.00

