Evaluate testing data (regression) - Lasso

Andrew Chang

2018-12-05

Contents

Labels: os_time

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

0. Load Data

```
## 300 of samples were used

## 101 of full features

## 2 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
```
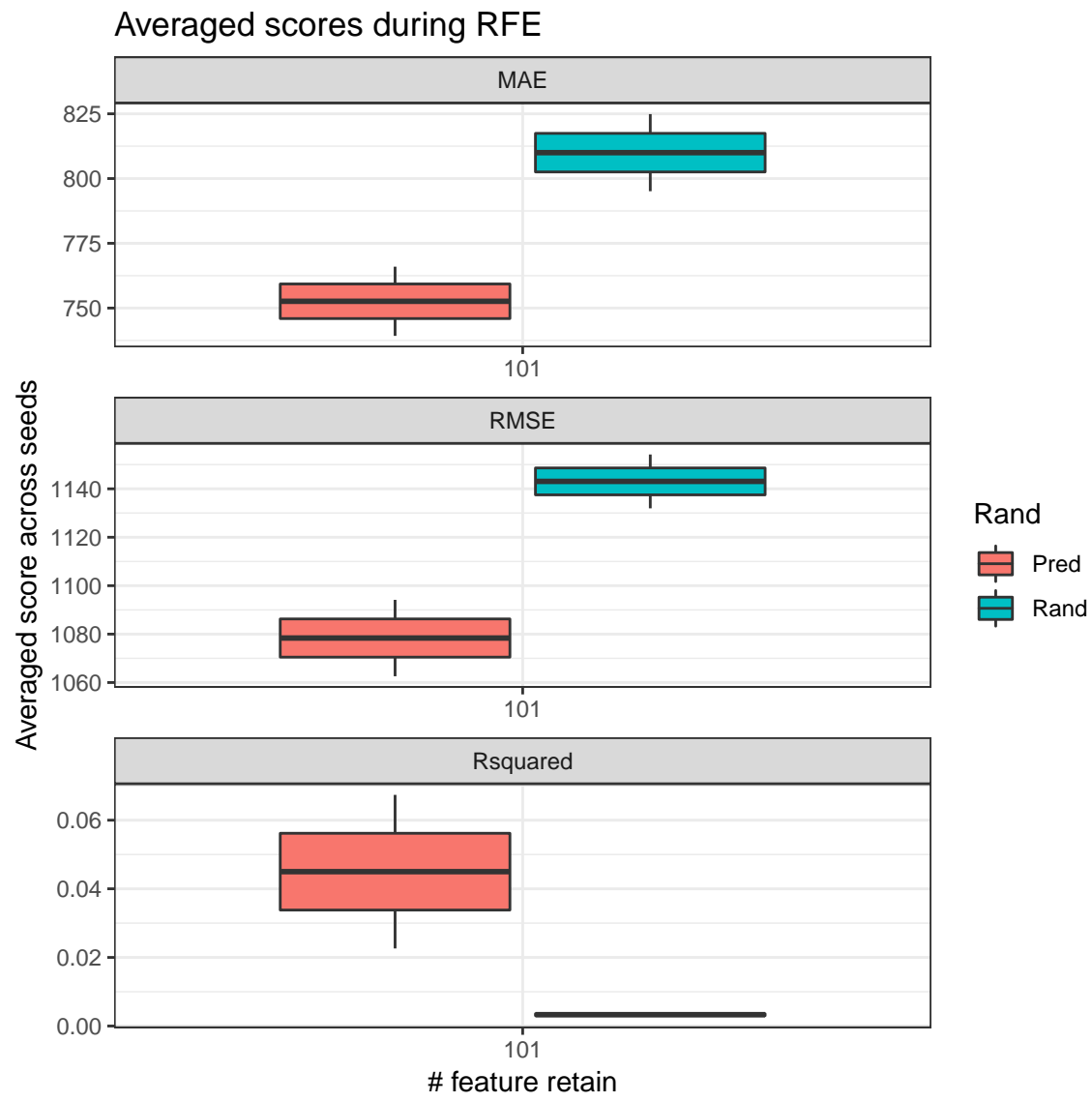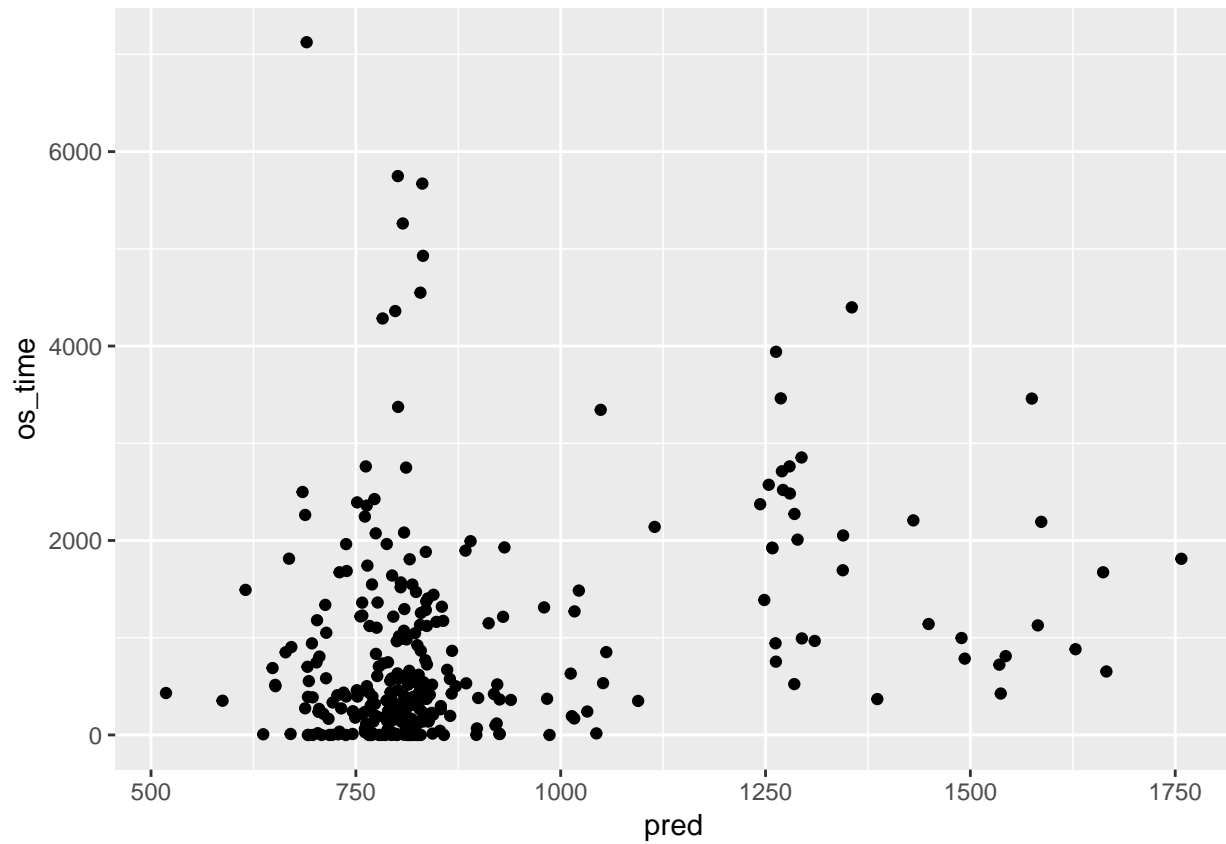
1. Scores

`average = T`: scores are based on pre-validation (combine predictions from all CVs per job and then calculate a single score per job).

`average = F`: report all the scores from each CV across entire jobs.

Averaged scores during RFE
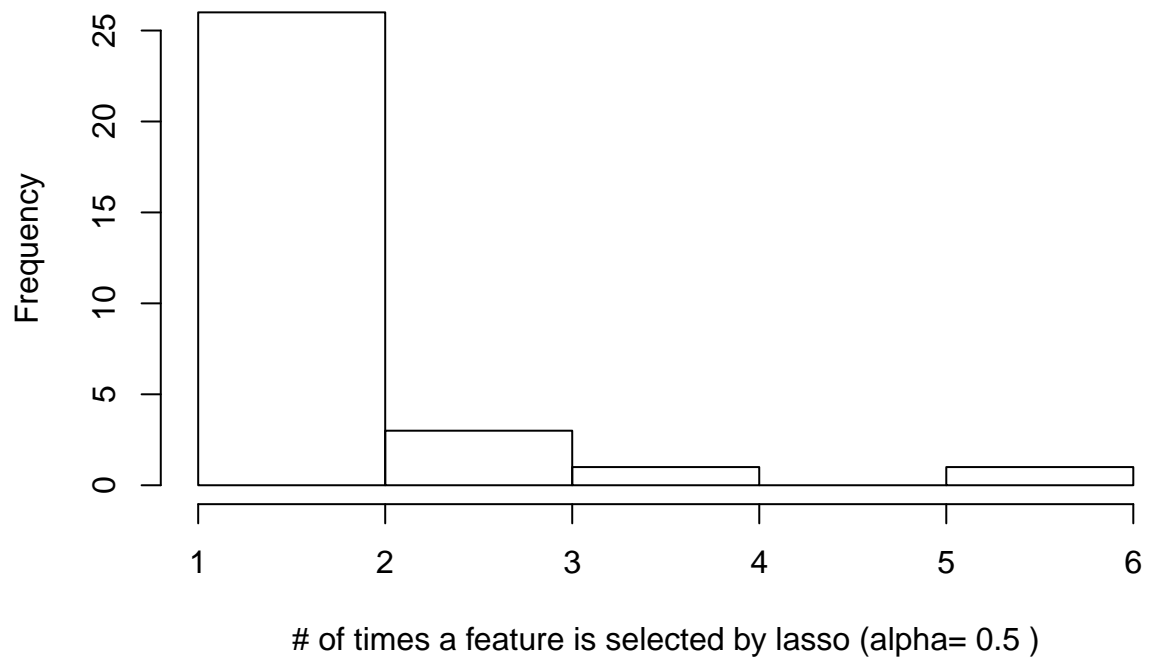
correlation


```
## # A tibble: 1 x 2
##   cor.avg cor.sdt
##     <dbl>   <dbl>
## 1   0.205  0.0772
```

2. Important Features

For Lasso, we calculate how many times a given features is being used in all the runs.

distribution across 2 seed x 3 CV

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

most used features