

Evaluate testing data (survival) - lasso

John Doe

2019-01-30

Contents

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

Labels: os_time

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

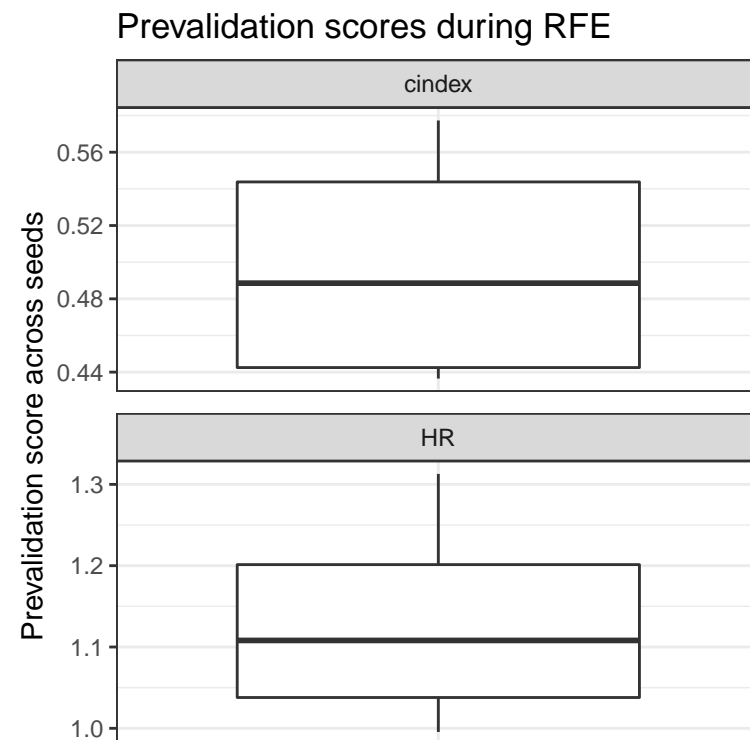
0. Load Data

300 of samples were used

100 of full features

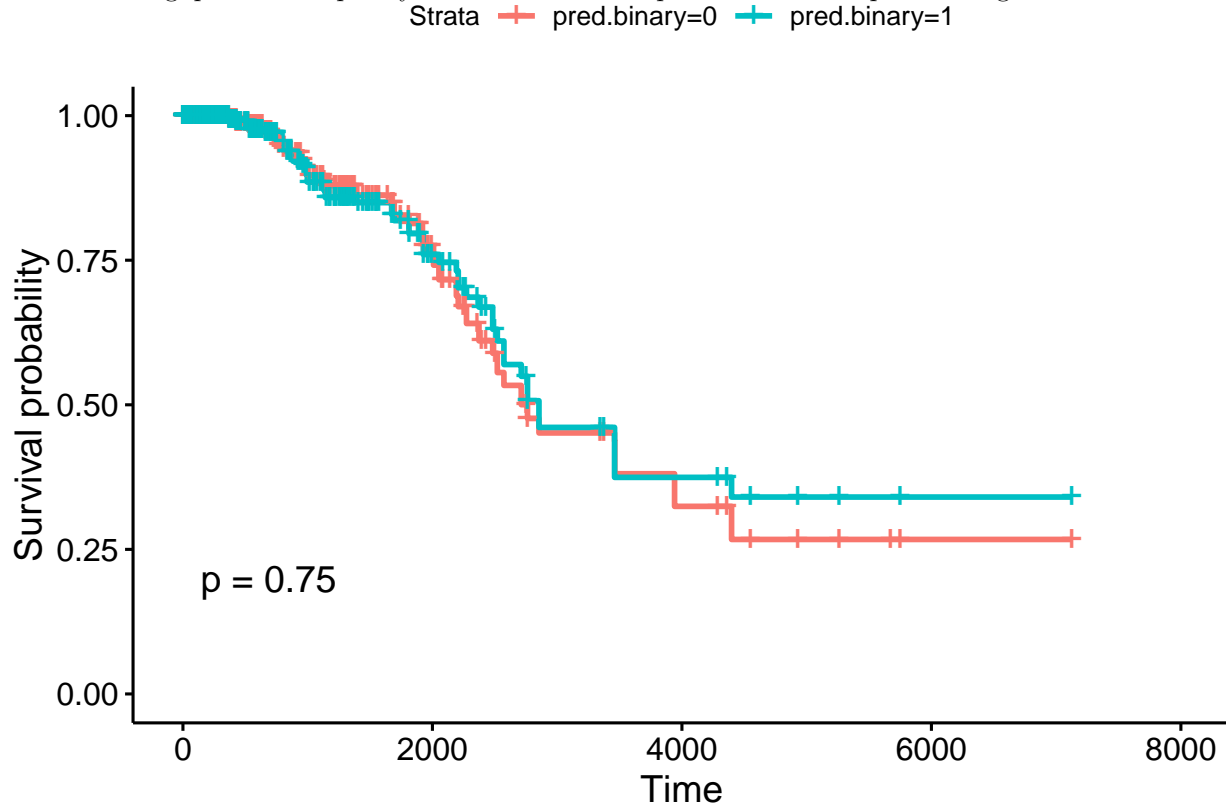
4 runs, each run contains 3 CVs.

1. Scores



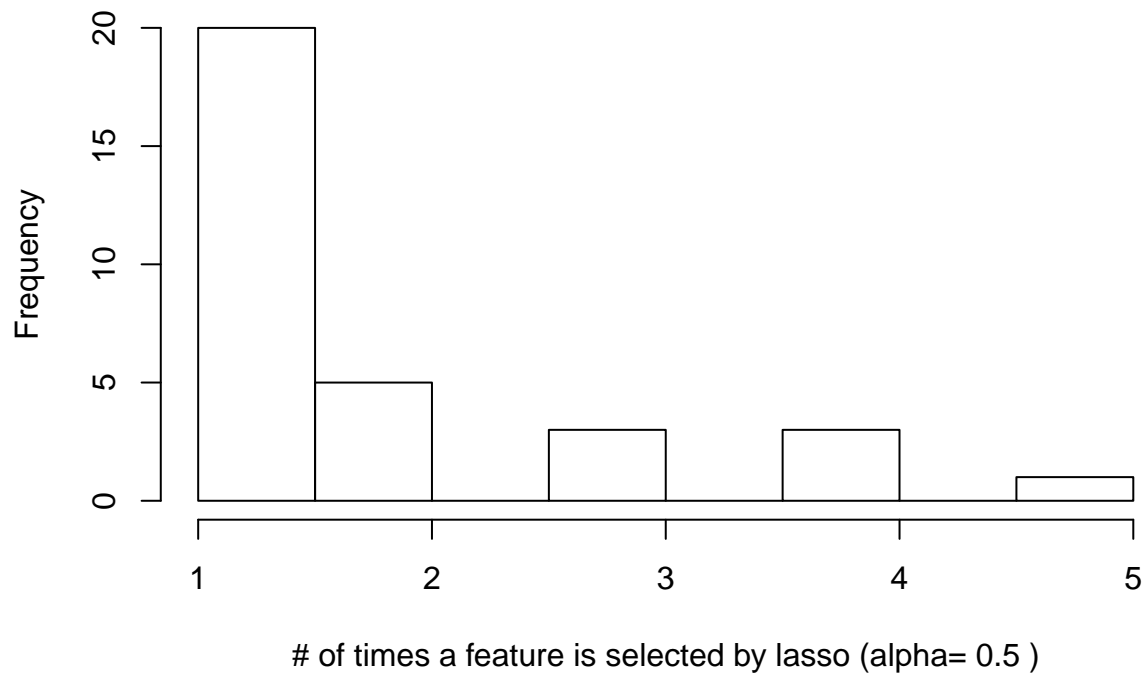
Note for the **HR plot**: A HR value (per seed) is calculated by comparing the survival time between 'long' and 'short' survivors. These two group is defined by splitting samples based on *median* predicted risk score; group_0 is predicted risk scores > median, which can be viewed as 'short survivors'. On the other hand, group_1 can be viewed as 'long survivors'. If the prediction is reasonable, the hazard ratio of group_1/group_0 should be < 1. The actual function used in calculating HR is `coxph(Surv(time, status) ~ group.binary, df)`.

The following plot is to quickly see how well the prediction can separate long and short survivor.

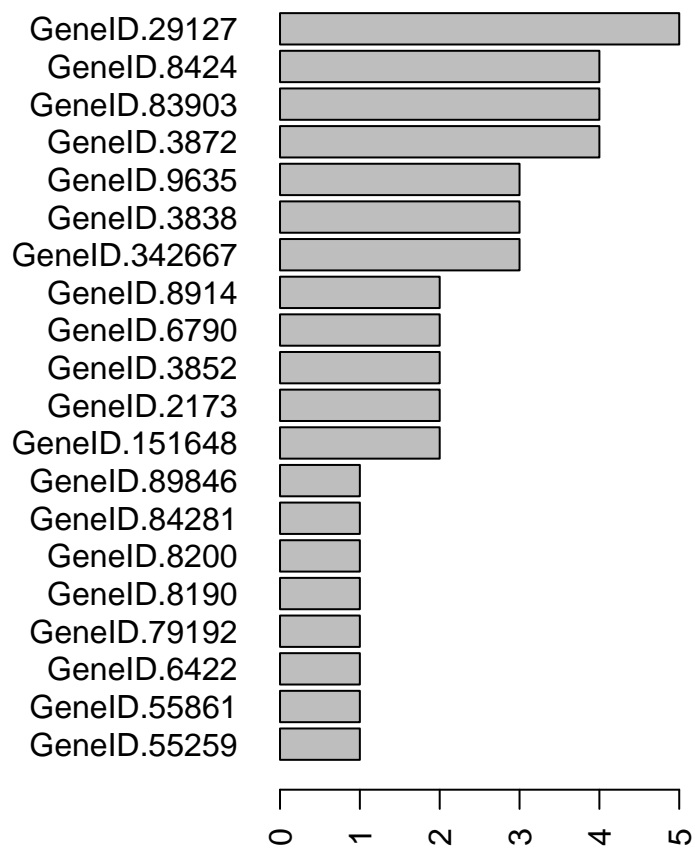


2. Important Features

distribution across 4 seed x 3 CV

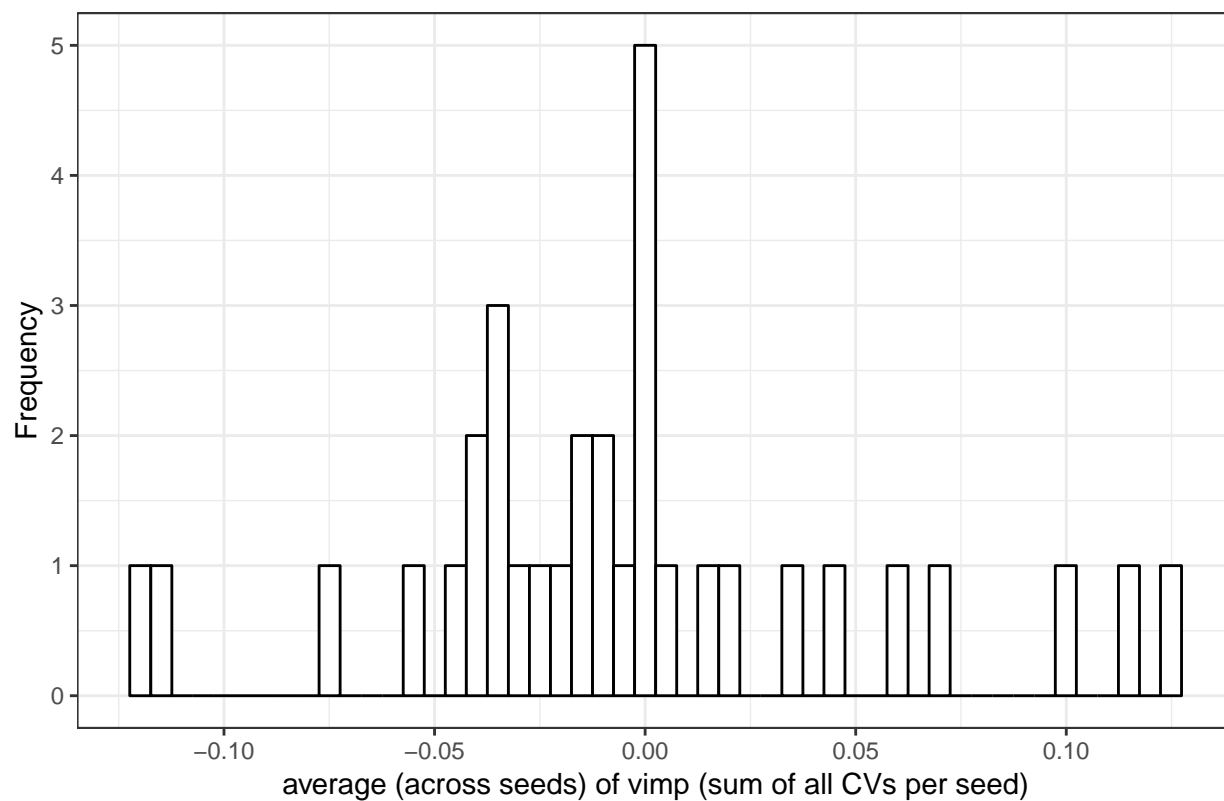


Number of times a feature is used

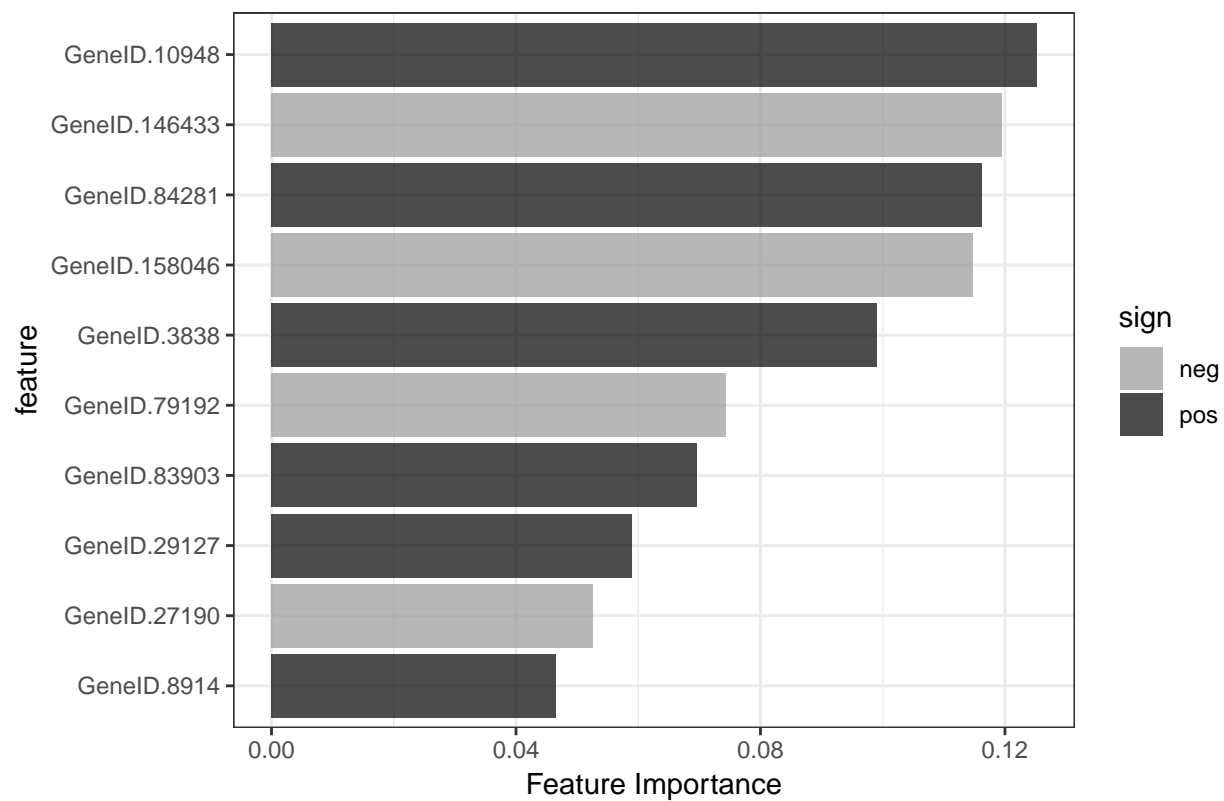


(currently only Lasso has this graph)

Distribution of Feature Coefficient



Top Features



Heatmap of top 20 important features

