Evaluate testing data (survival) - lasso

EVE W.

2020-04-19

Contents

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

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

0. Load Data

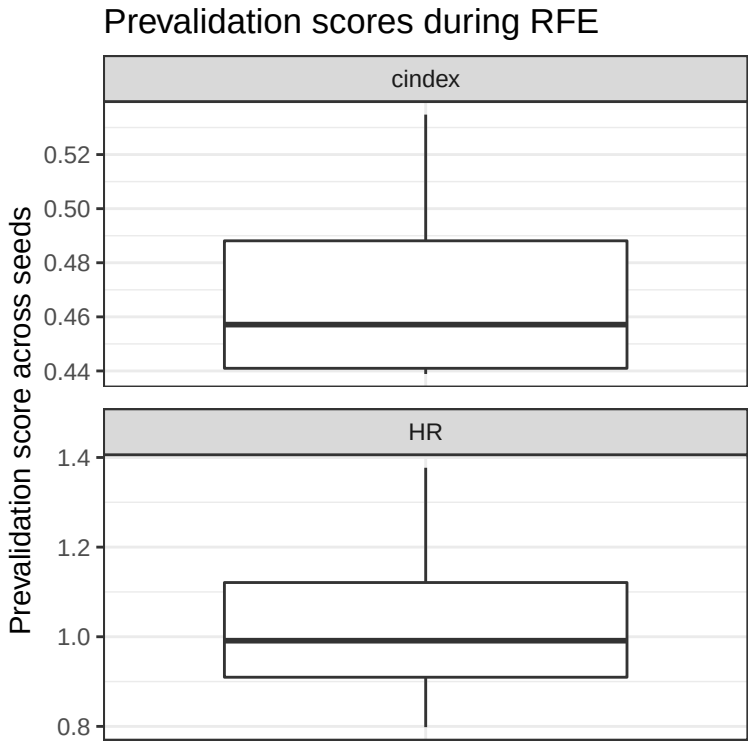
```
300 of samples were used
```

```
100 of full features
```

```
4 runs, each run contains 3 CVs.
```

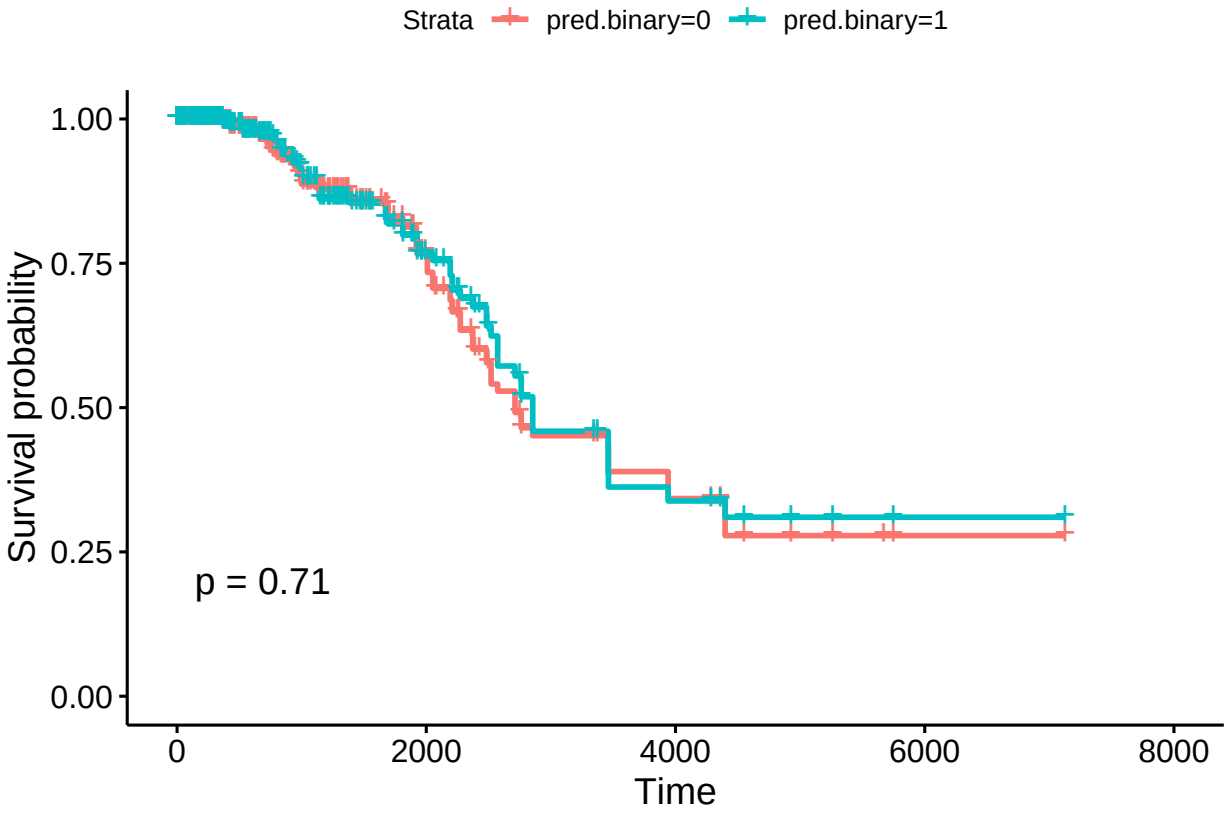
run with lasso.r.

1. Scores

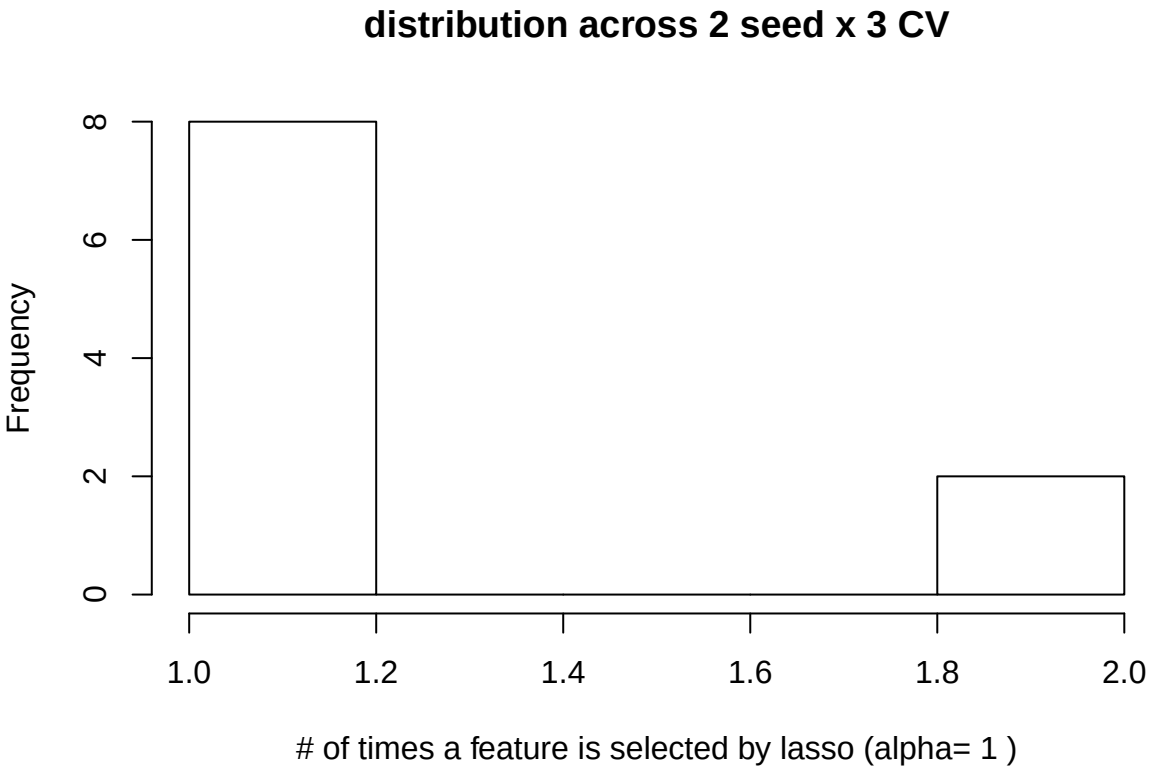


Note for the **HR plot**: A HR value (per seed) is calculated by comparing the survival time between 'long' and 'short' survivors. These two group is defined by splitting samples based on *median* predicted risk score; group_0 is predicted risk scores > median, which can be viewed as 'short survivors'. 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.



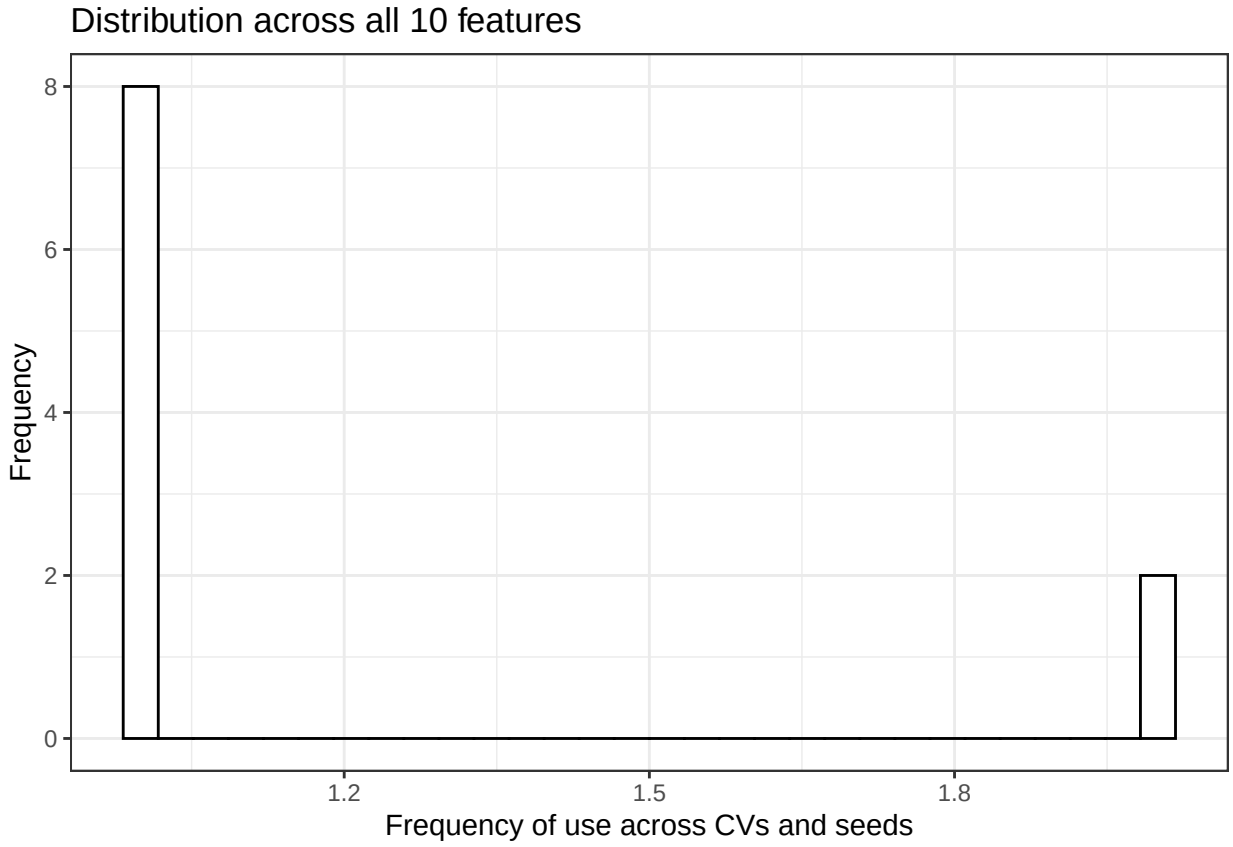
2. Important Features

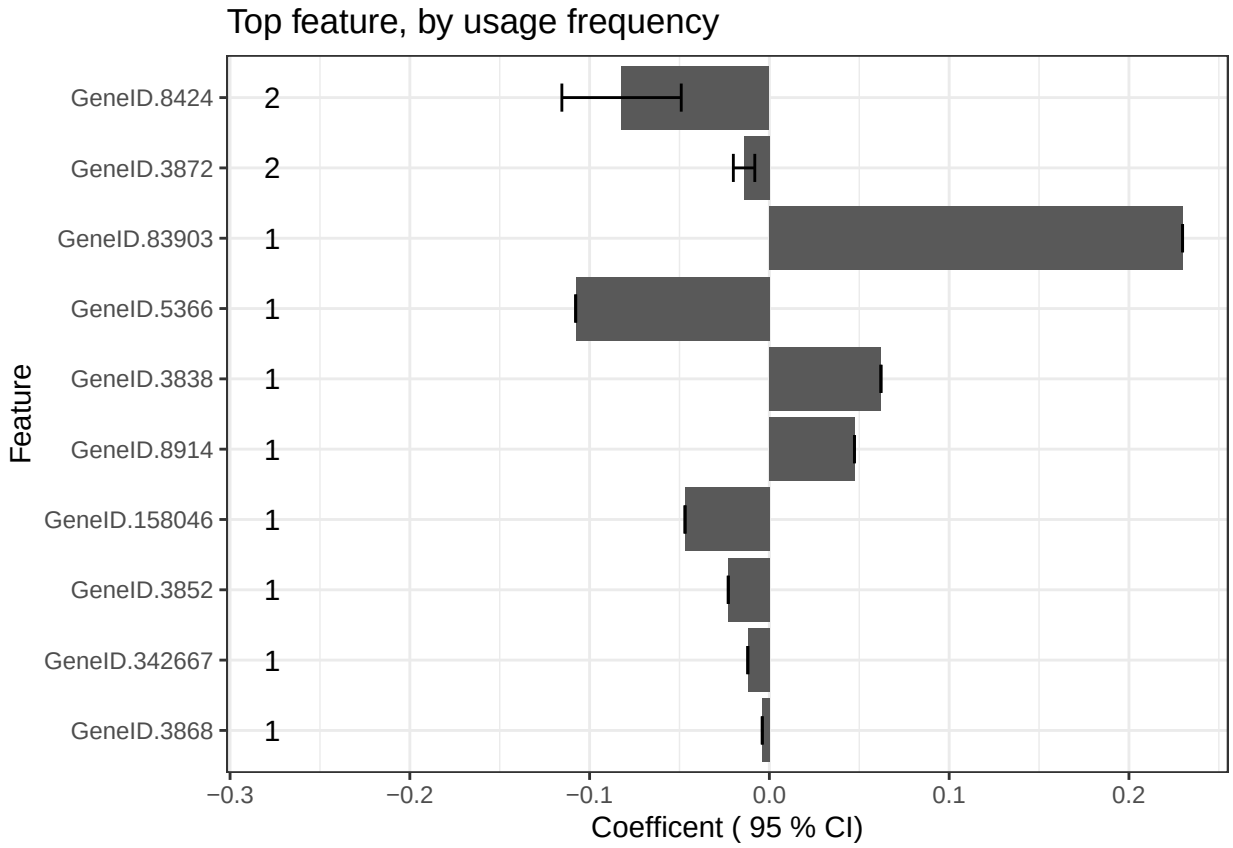


```
## [1] "there are 10 unique features used from the 100 feature set"
## [1] "summary of number of features used in each run under 2 seeds and 3 CVs"
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       1       1       2       2       3       3
```

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




Top feature, by usage frequency
Feature
GeneID.8424
GeneID.3872
GeneID.83903
GeneID.5366
GeneID.3838
GeneID.8914
GeneID.158046
GeneID.3852
GeneID.342667
GeneID.3868
2
2
1
1
1
1
1
1
1
1
−0.3
−0.2
−0.1
0.0
0.1
0.2
Coefficent (95 % CI)