

Evaluate testing data (survival) - rfsrc

EVE W.

2020-06-11

Contents

0. Load Data	1
1. Scores	2
2. Important Features	5
marginal effects	6

Label: os_time

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

0. Load Data

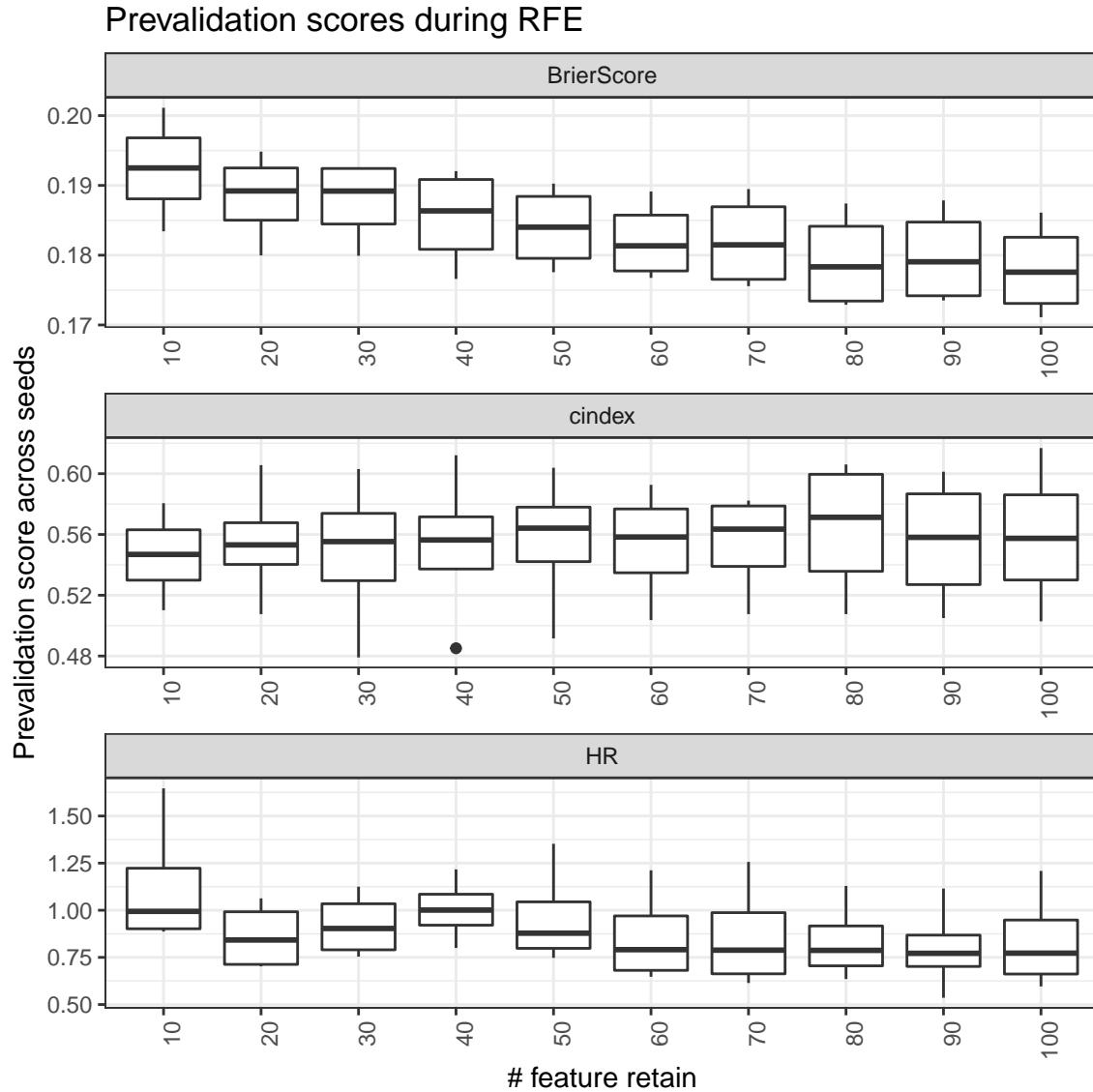
276 of samples were used

100 of full features

4 runs, each run contains 3 CVs.

run with rfeSRCC.r.

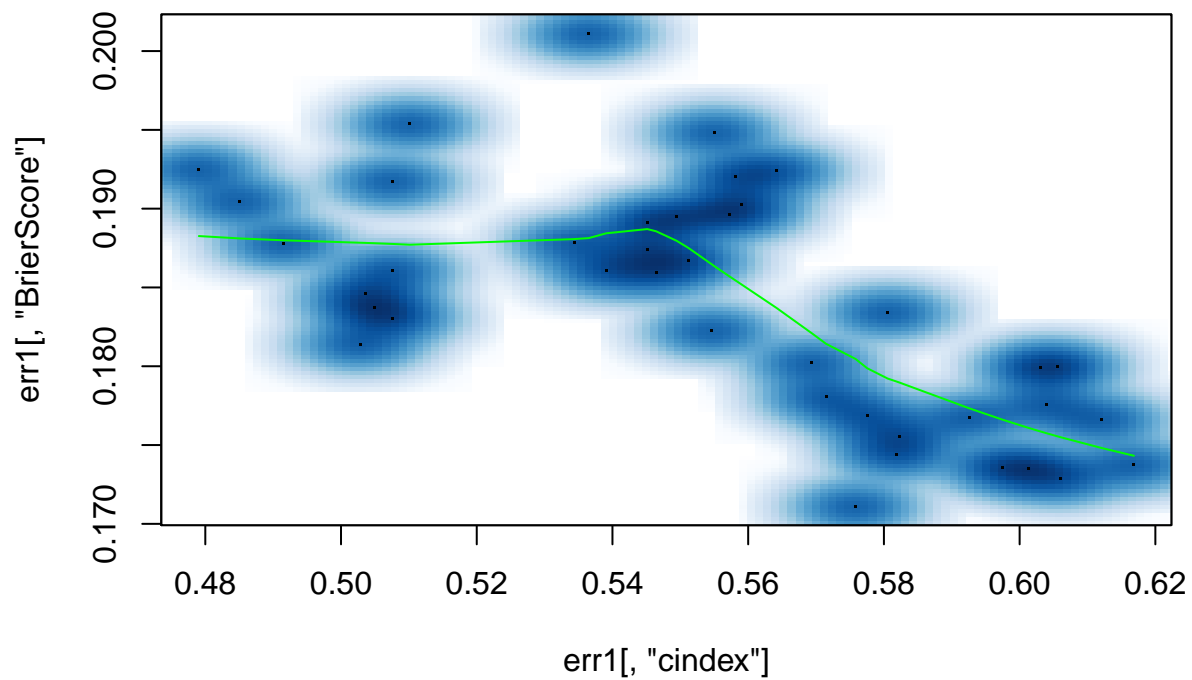
1. Scores



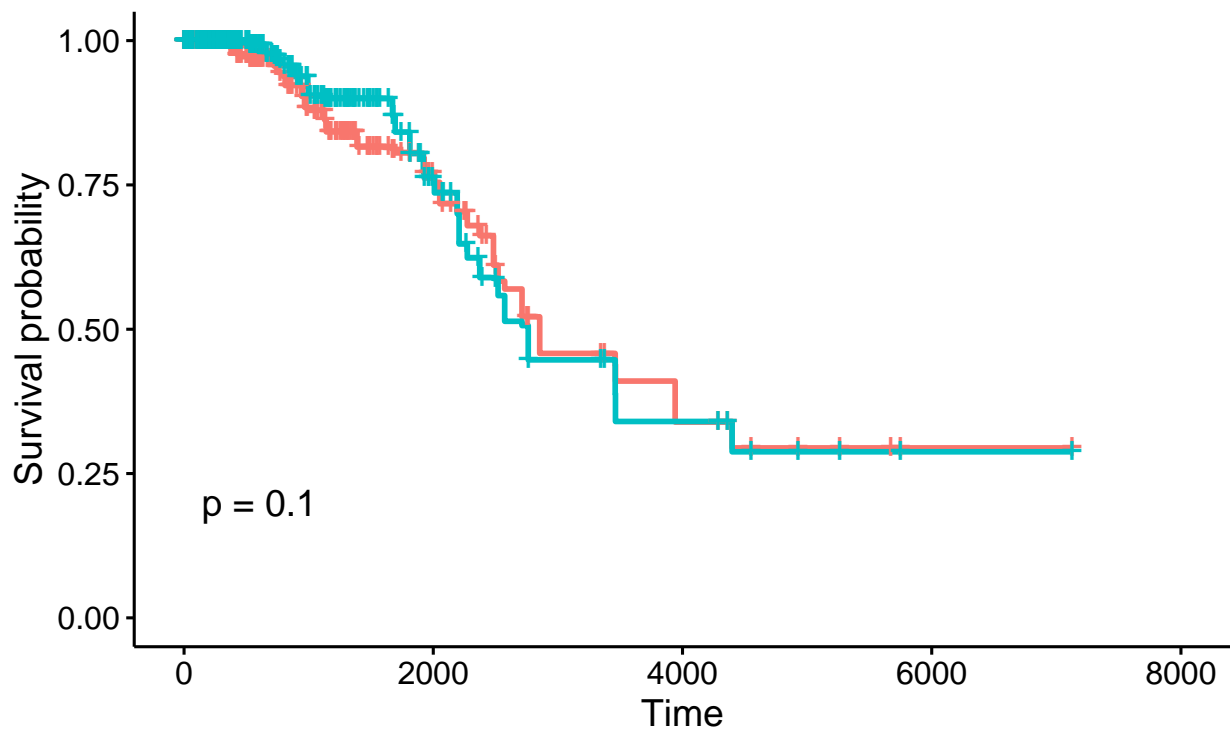
Note for the **HR plot**: A HR value (per seed) is calculated by comparing the survival time between 'long' and 'short' survivors. These two group is defined by splitting samples based on *median* predicted risk score; group_0 is predicted risk scores > median, which can be viewed as 'short survivors'.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

metrics	size.max	median.max	size.min	median.min
BrierScore	10	0.193	100	0.178
cindex	80	0.571	10	0.547
HR	40	1.001	90	0.772

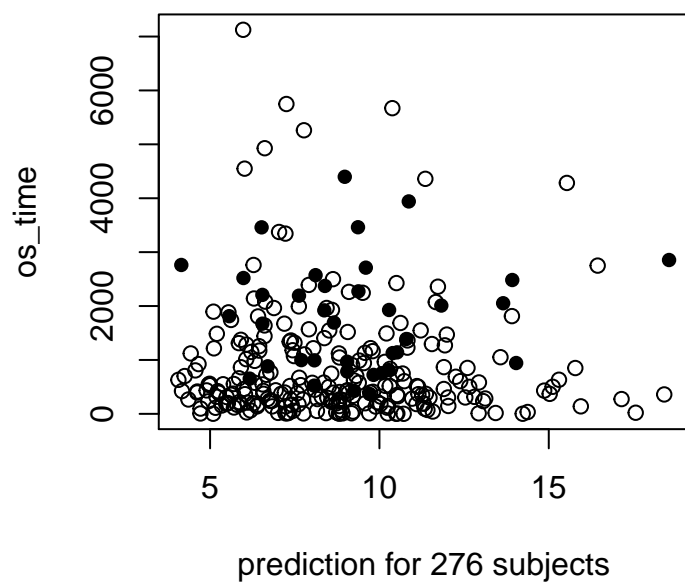
pearson corr: -0.65



The following plot is to quickly see how well the prediction can separate long and short survivor.
Strata + pred.binary=0 + pred.binary=1

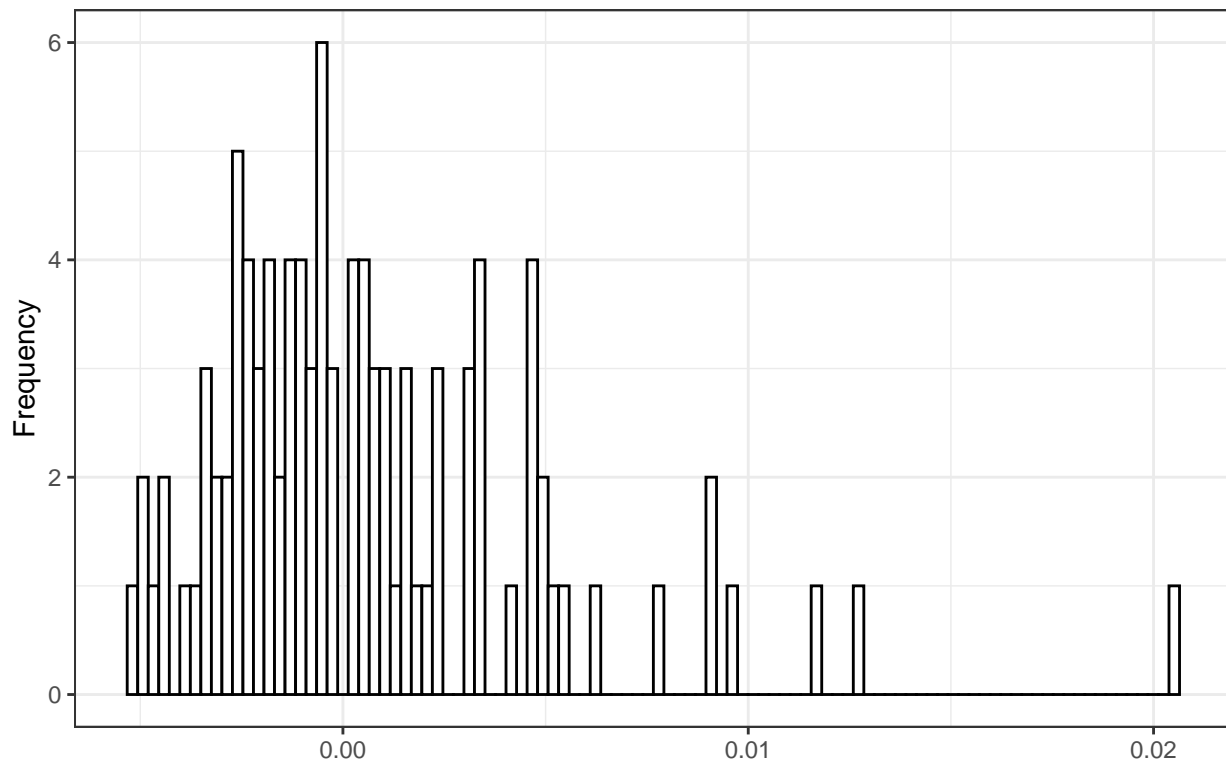


prediction under seed 1001 at full feature s



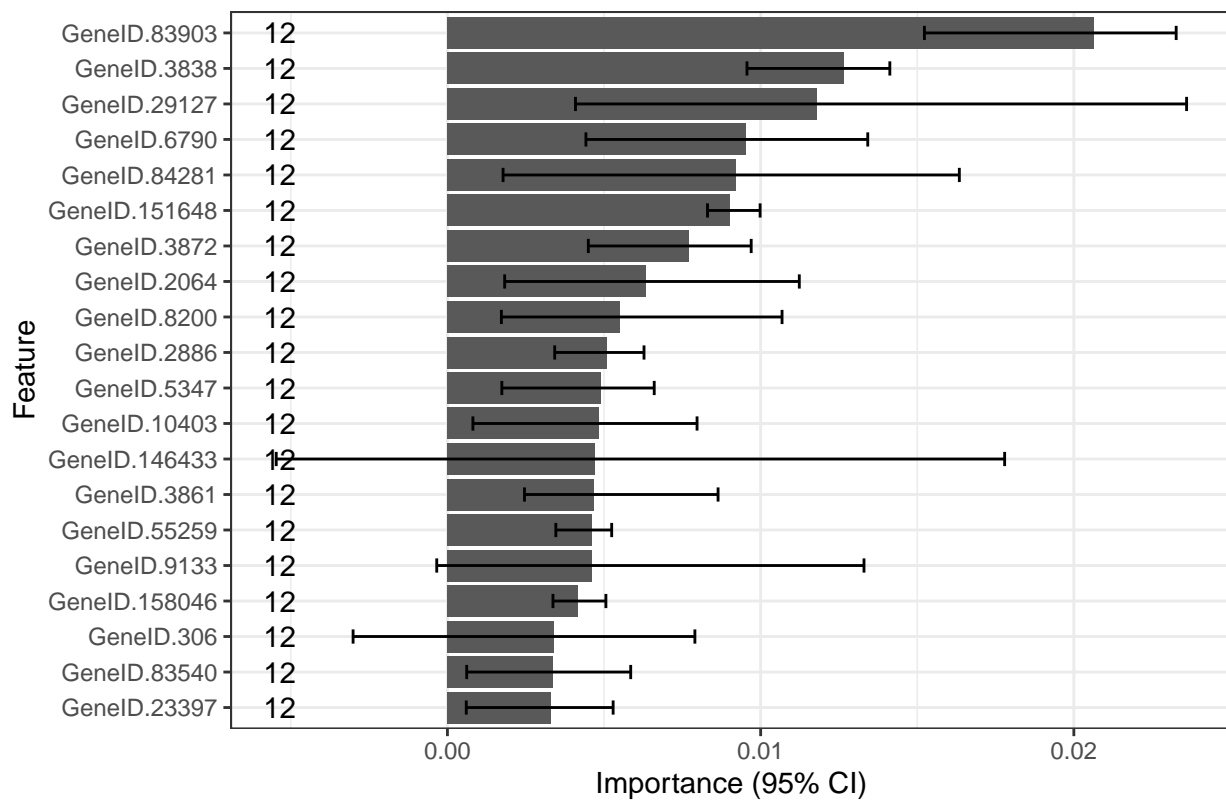
2. Important Features

From 100 feature step based on vimp



Average (across seeds) of vimp (sum of all CVs per seed)

Feature importance by VIMP magnitude



marginal effects

