Evaluate testing data (survival) - rfsrc

EVE W.

2019-06-01

Contents

Label: os_time

```
## user input
project_home <- "~/EVE/examples"
project_name <- "rfsrc_outCV_test"
```
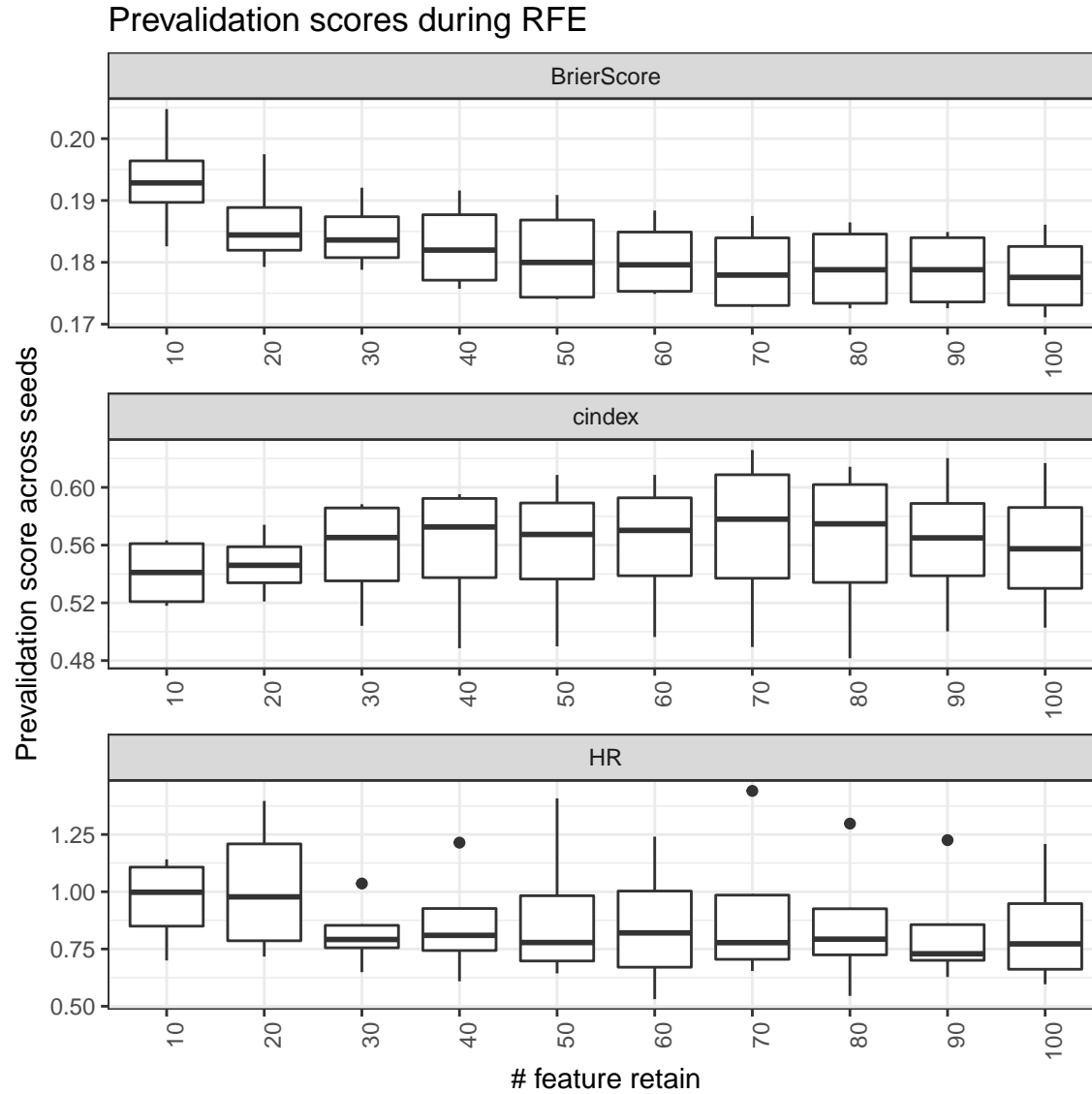
0. Load Data

```
276 of samples were used
```

```
100 of full features
```

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

run with rfeSRCC.r.
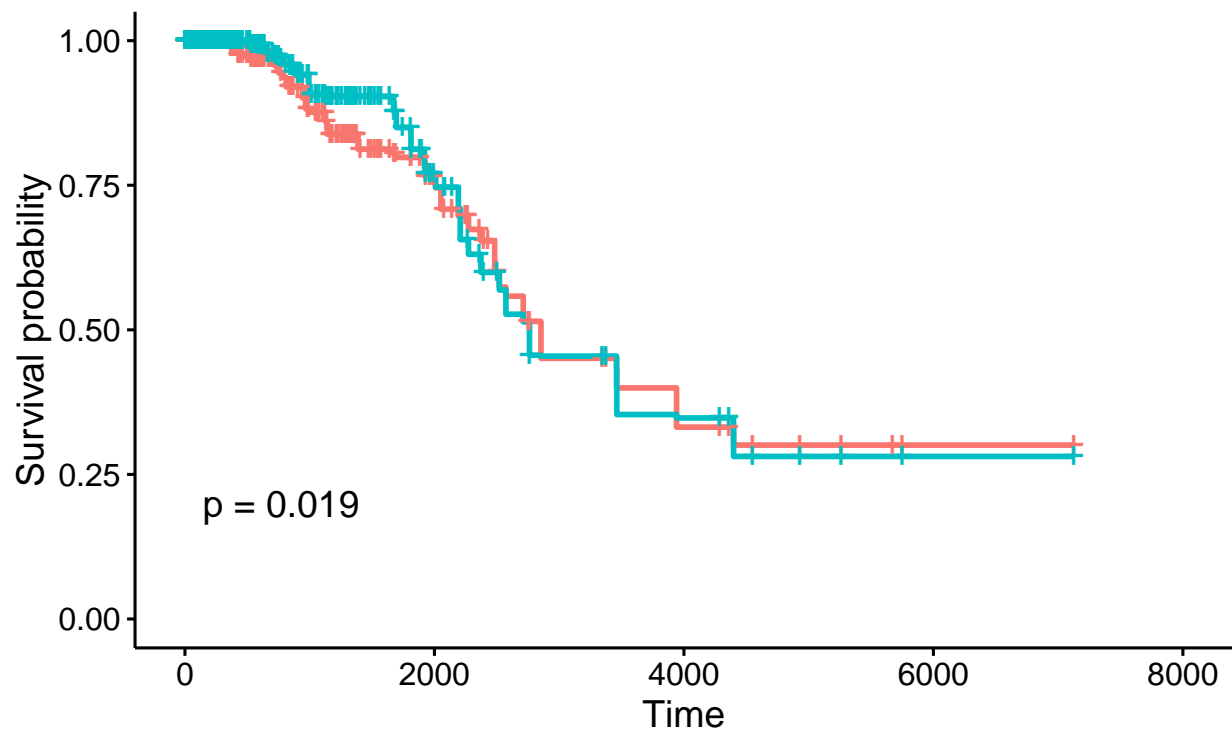
1. Scores

Prevalidation scores during RFE

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	70	0.578	10	0.541
HR	10	0.998	90	0.729

pearson corr: −0.63
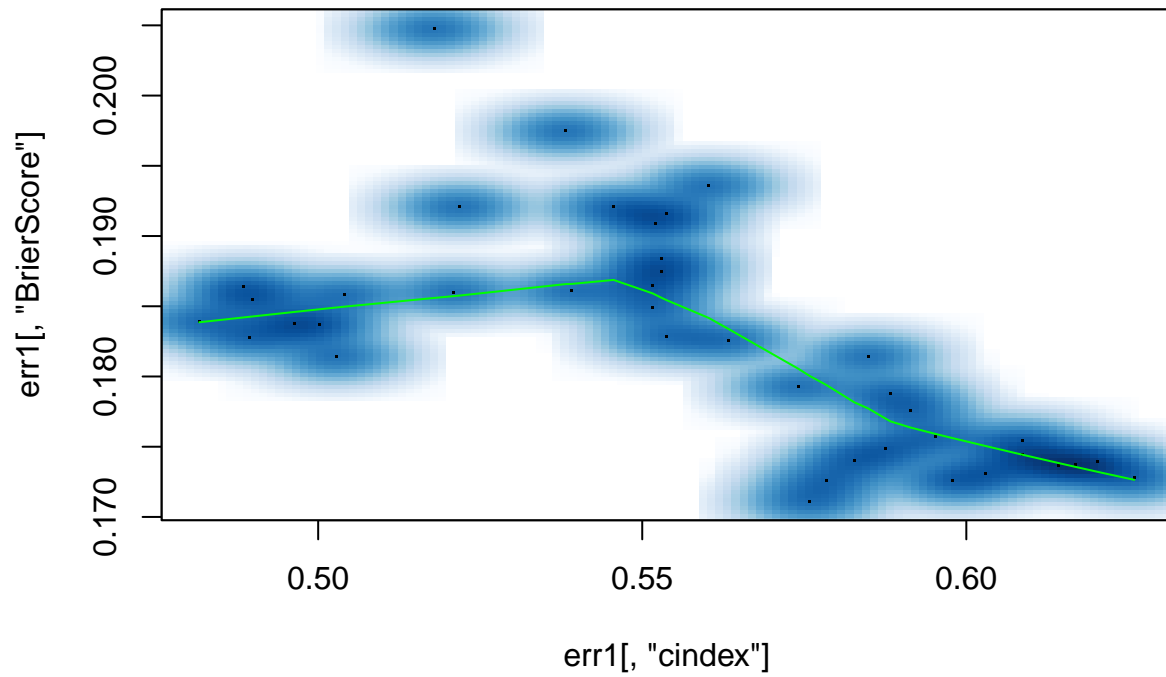
The following plot is to quickly see how well the prediction can separate long and short survivor.

Strata ┼ pred.binary=0 ┼ pred.binary=1

p = 0.019

2. Important Features

From 100 feature step based on vimp

Feature importance by usage frequency

Heatmap of top 20 important features