

Evaluate testing data (survival) - rfsrc

Andrew Chang

2019-01-09

Contents

0. Load Data	1
1. Scores	2
2. Important Features	4

Label: os_time

```
## user input
project_home <- "~/EVE/tests"
project_name <- "rfsrc_outCV_test"
## end of user input
```

0. Load Data

Warning: `data_frame()` is deprecated, use `tibble()`.

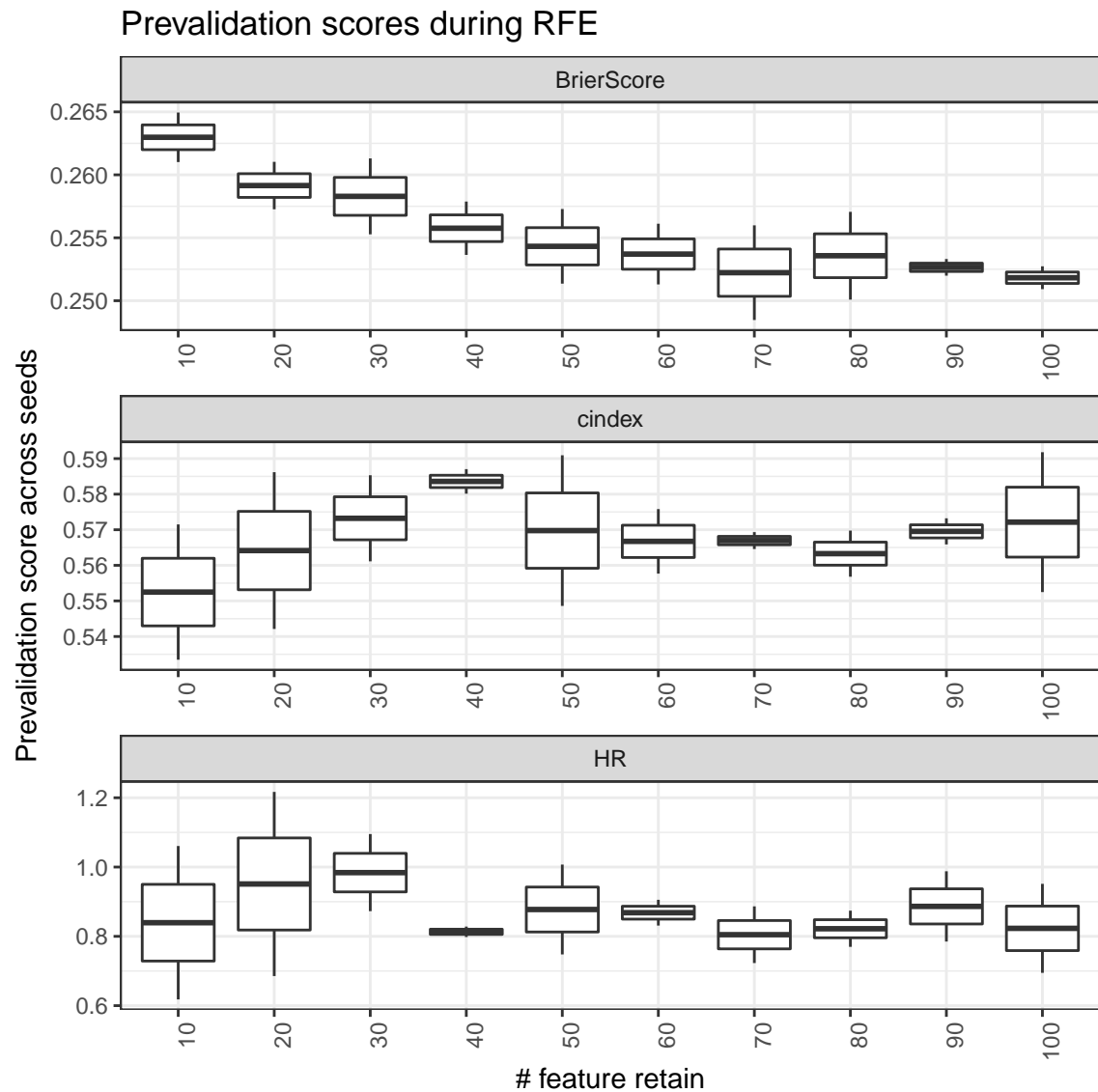
This warning is displayed once per session.

300 of samples were used

100 of full features

2 runs, each run contains 3 CVs.

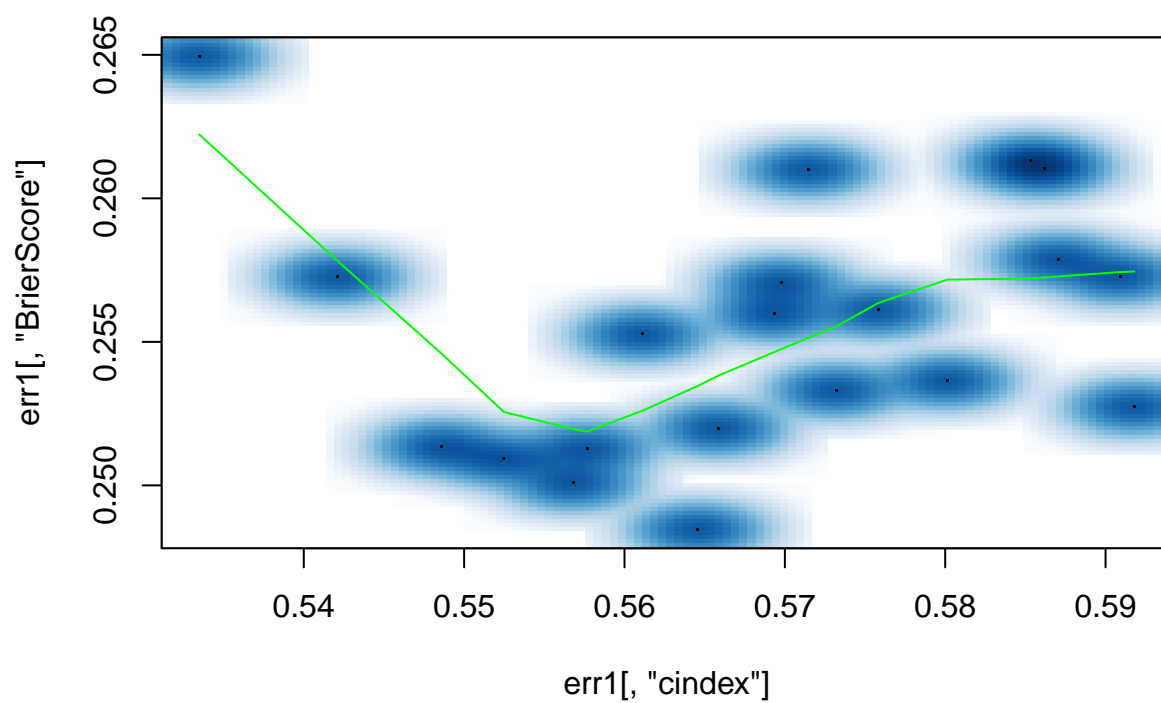
1. Scores



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)`.

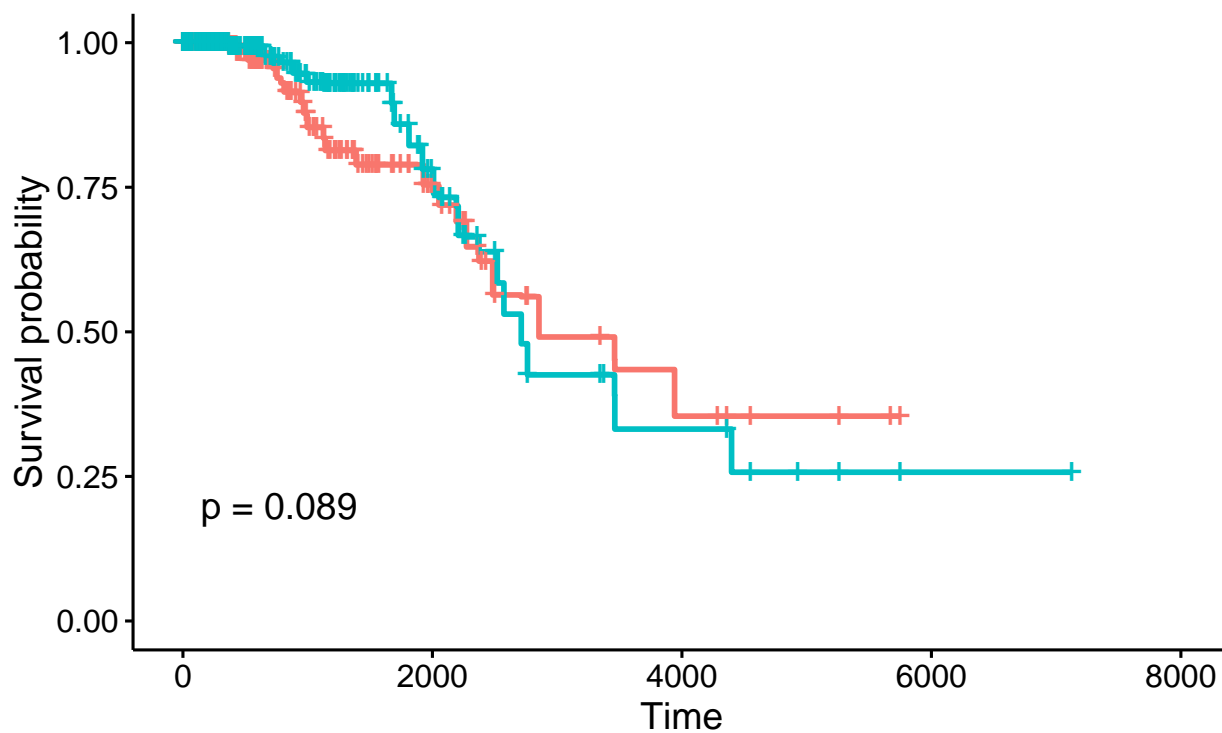
```
err1 <- data.frame(plt$df.scores)
smoothScatter(err1[, 'cindex'], err1[, 'BrierScore'],
               main=paste('pearson corr:', round(cor(err1[, 'cindex'], err1[, 'BrierScore']), 2) ))
lines(lowess(err1[, 'cindex'], err1[, 'BrierScore']), col = 'green')
```

pearson corr: 0.1



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

Strata + pred.binary=0 + pred.binary=1



2. Important Features

