Evaluate testing data (survival) - rfsrc

John Doe

2019-01-30

Contents

Label: os_time

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

0. Load Data

```
300 of samples were used
```
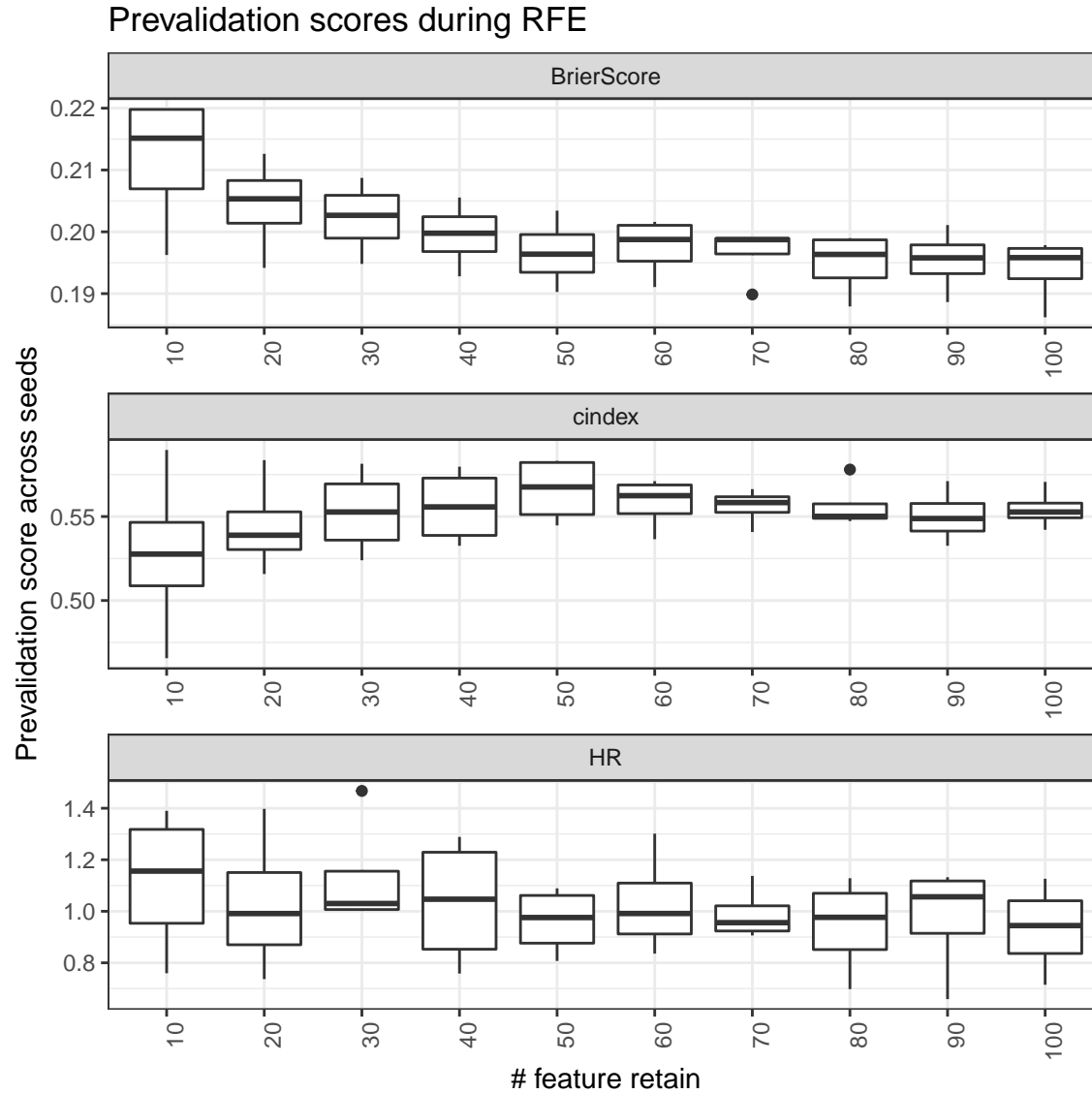
```
100 of full features
```

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

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	0.215	50	0.196
cindex	50	0.568	10	0.528
HR	10	1.156	100	0.944

pearson corr: −0.73
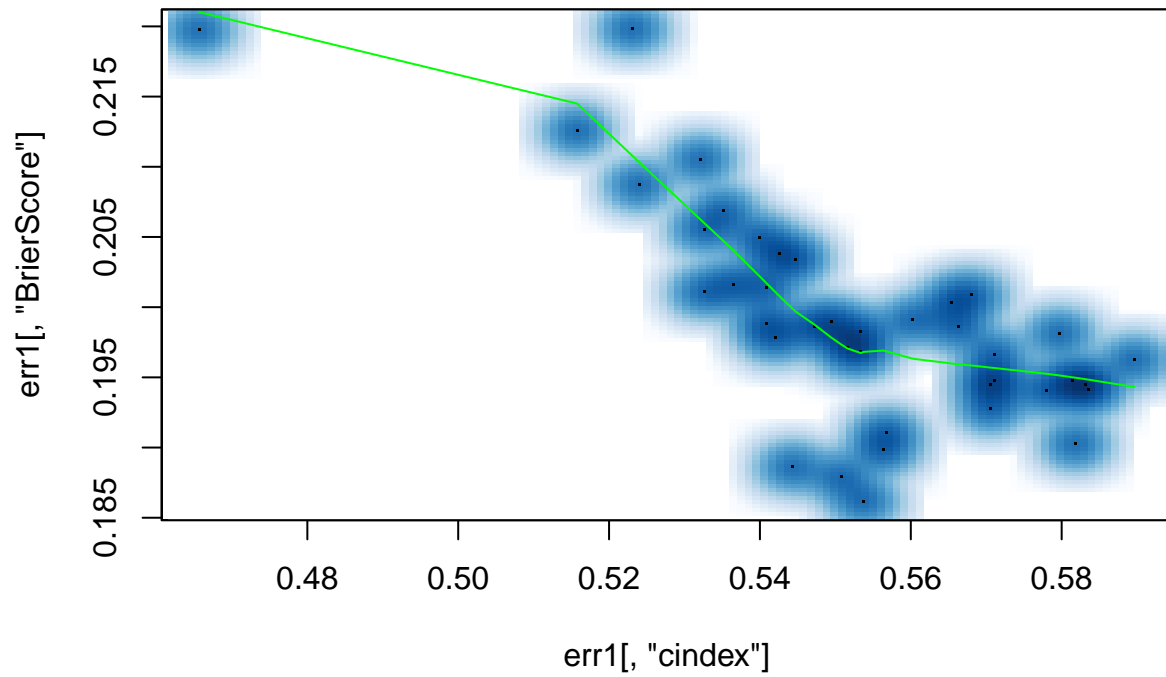
The following plot is to quickly see how well the prediction can separate long and short survivor.
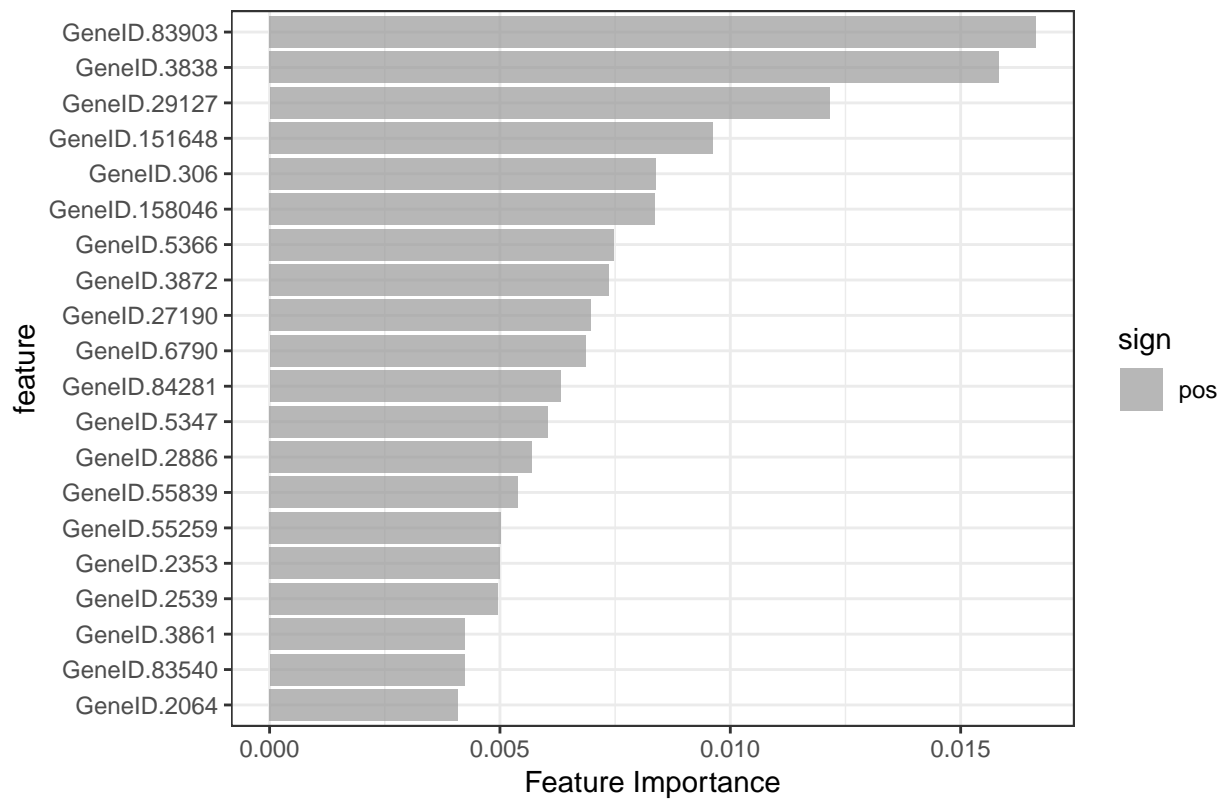
Strata ┼ pred.binary=0 ┼ pred.binary=1

p = 0.89
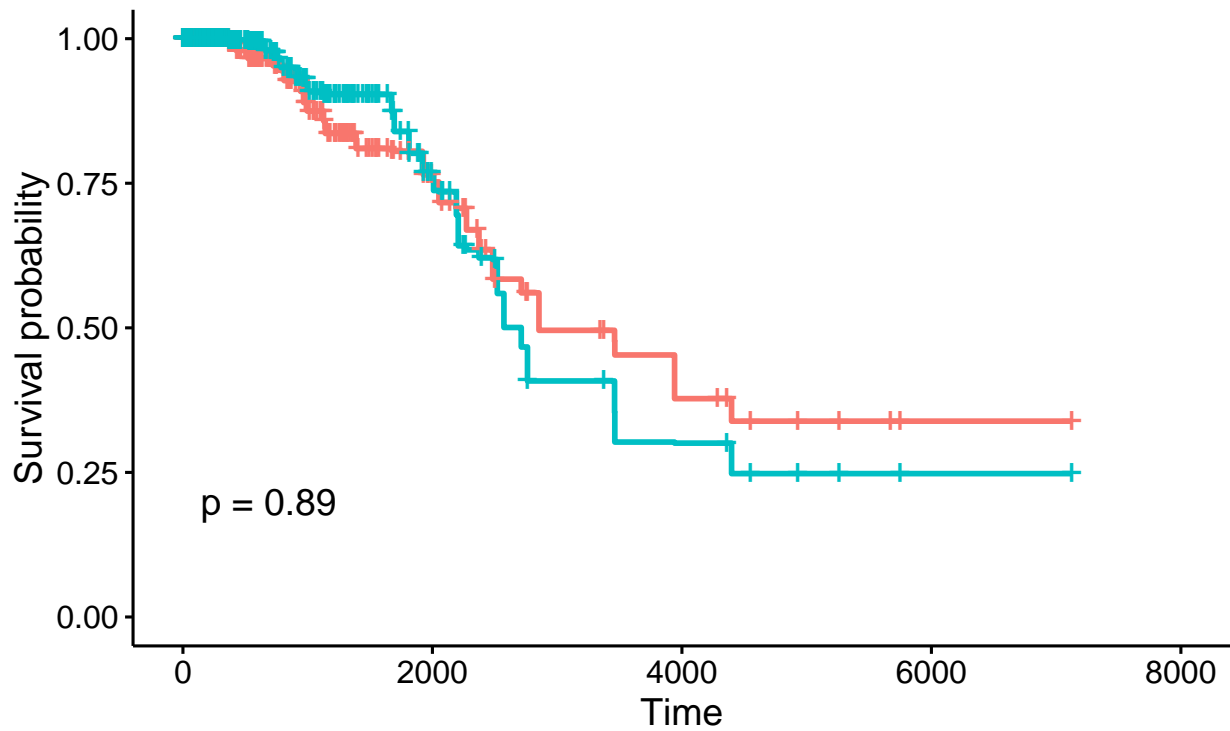
2. Important Features

with 100 features based on vimp

average (across seeds) of vimp (sum of all CVs per seed)
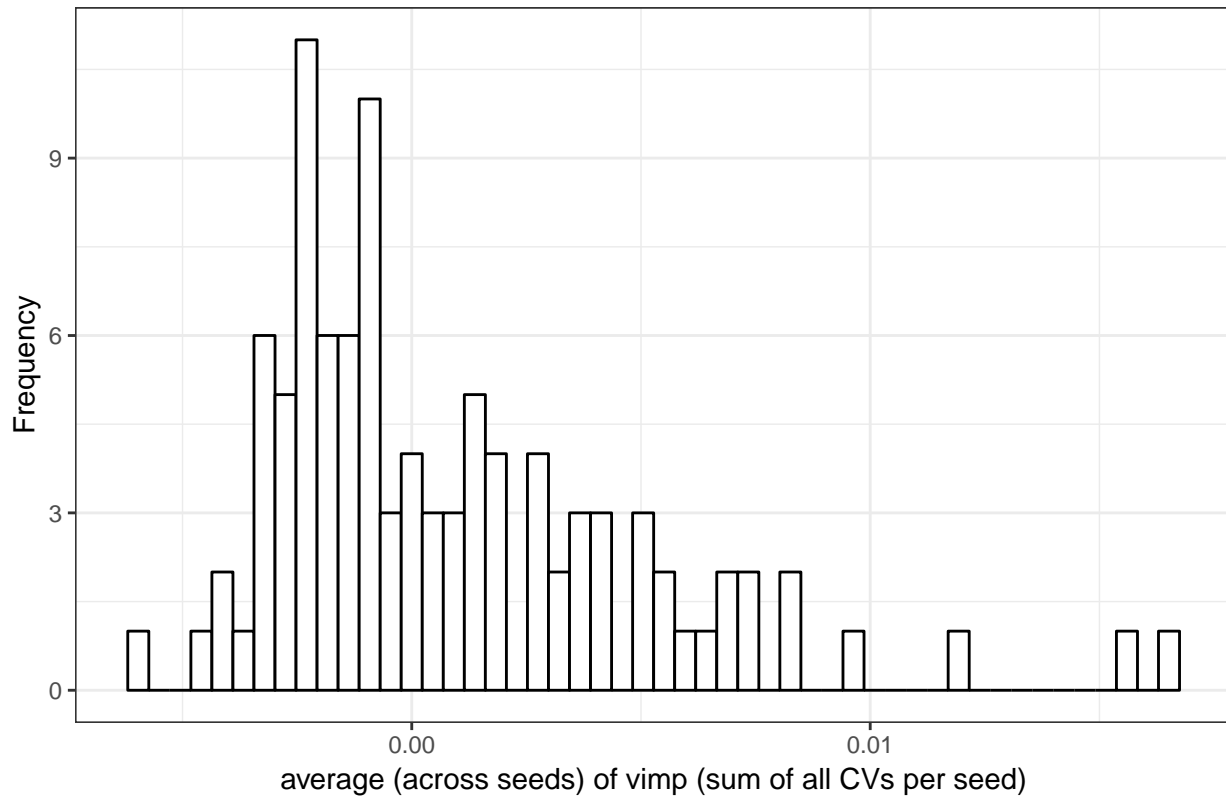
Top 20 features at 100 feature set

Heatmap of top 20 important features