Simulated examples for the survival ensemble methods

Select data type

W500

Time point for event prediction:

5

Random seed for calibration and validation

42

K_Outer loop CV (for validation)

3

K_ inner CV folds (model tuning)

3

Simulated data: random seed (generation):

4242

Sample size:

150

Observation time

5

Expected event prevalence by study end

0.5

Expected drop out rate

0.3

Custom data: path to data file

~/Desktop/Study_KCL/PhD P

Predictors to use in the model

"baseline_age_", "genderdun

Time variable name

time

Event indicator variable name

event

Sample statistics | CoxPH | SRF | Ens1: CoxPH->SRF | Ens2: CoxPH in clusters | Ens3: extended CoxPH | Summary | Conclusions

Internally cross-validated results:

Show 10 entries Search:

	AUCROC	BS	BS_scaled	C_score	Calib_slope	Calib_alpha	T
test	0.8373	0.2319	0.1816	0.7821	1.7767	0.0753	5
train	0.9039	0.1788	0.3731	0.8238	2.1936	0.0637	5

Showing 1 to 2 of 2 entries Previous 1 Next

Internally cross-validated Test results for each CV fold:

Show 10 entries Search:

	AUCROC	BS	BS_scaled	C_score	Calib_slope	Calib_alpha	T
test.1	0.8723	0.3215	-0.0934	0.8065	2.3465	-0.0562	5
test.2	0.8136	0.1877	0.3262	0.7917	1.4341	0.1131	5
test.3	0.826	0.1865	0.312	0.7482	1.5495	0.1691	5

Showing 1 to 3 of 3 entries Previous 1 Next

```
$test
  T    AUCROC        BS  BS_scaled   C_score Calib_slope
1 5 0.8722601 0.3214817 -0.0933980 0.8064753    2.346536
2 5 0.8135598 0.1877153  0.3261991 0.7917345    1.434069
3 5 0.8260365 0.1864932  0.3119813 0.7482374    1.549465
  Calib_alpha test cv_n
1  -0.0562467    1    1
2   0.1130697    1    2
3   0.1691393    1    3

$train
  T    AUCROC        BS BS_scaled   C_score Calib_slope
1 5 0.8762396 0.2035782 0.2986905 0.8083743    2.158375
2 5 0.9183884 0.1696934 0.3996468 0.8143104    2.112169
3 5 0.9169287 0.1632106 0.4210484 0.8487737    2.310226
  Calib_alpha test cv_n
1  0.05931673    0    1
2  0.06713708    0    2
3  0.06466375    0    3

$testaverage
          T      AUCROC          BS   BS_scaled
 5.00000000  0.83728546  0.23189670  0.18159412
    C_score Calib_slope Calib_alpha        test
 0.78214907  1.77668994  0.07532078  1.00000000

$trainaverage
          T      AUCROC          BS   BS_scaled
 5.00000000  0.90385225  0.17882740  0.37312860
    C_score Calib_slope Calib_alpha        test
 0.82381948  2.19358982  0.06370585  0.00000000

$model_list
$model_list[[1]]
$model_list[[1]]$beststats
   mtry nodesize nodedepth time    AUCROC        BS
V1    3       25        50  3.3 0.8042584 0.1854946
   BS_scaled  C_score Calib_alpha Calib_slope
V1 0.2515693 0.747072  0.07469976    1.414087

$model_list[[1]]$allstats
    mtry nodesize nodedepth time    AUCROC        BS
V1     4       15        50  3.3 0.8036135 0.1827861
V2     4       20        50  3.3 0.8021343 0.1837010
V3     4       25        50  3.3 0.8050936 0.1828498
V4     4       30        50  3.3 0.8040434 0.1838655
V11    3       25        50  3.3 0.8042584 0.1854946
V21    5       25        50  3.3 0.7956085 0.1858166
V31    7       25        50  3.3 0.7932647 0.1871167
V41   10       25        50  3.3 0.7908765 0.1878931
V5    15       25        50  3.3 0.7835970 0.1918826
    BS_scaled   C_score Calib_alpha Calib_slope
V1  0.2624972 0.7478766  0.07073499   1.1300125
V2  0.2588058 0.7493965  0.07188311   1.1877798
V3  0.2622401 0.7492177  0.07434891   1.2789207
```

```
V4  0.2581421 0.7472508  0.07595773   1.3479644
V11 0.2515693 0.7470720  0.07469976   1.4140874
V21 0.2502700 0.7474892  0.06849913   1.1332781
V31 0.2450242 0.7460885  0.06692381   1.0277256
V41 0.2418916 0.7424527  0.06683811   0.9110972
V5  0.2257949 0.7379228  0.06700300   0.8196784

$model_list[[1]]$model
                         Sample size: 333
                    Number of deaths: 148
                     Number of trees: 500
           Forest terminal node size: 25
       Average no. of terminal nodes: 10.412
No. of variables tried at each split: 3
              Total no. of variables: 17
       Resampling used to grow trees: swor
    Resample size used to grow trees: 210
                            Analysis: RSF
                              Family: surv
                    Splitting rule: logrank *random*
       Number of random split points: 50
                          (OOB) CRPS: 0.17290317
   (OOB) Requested performance error: 0.25534572


$model_list[[1]]$vimp10
          age   cox_predict           chf           bmi
 5.269667e-02  5.000389e-02  1.545431e-02  1.022894e-02
           hr           los           sho        diasbp
 6.643624e-03  3.182544e-03  9.647674e-04  6.115573e-04
       gender         miord           afb        mitype
 3.909241e-04  3.673733e-04  5.937008e-05  5.737802e-05
        y1997           cvd         y1999
 1.582724e-05 -1.293352e-04 -1.377054e-04

$model_list[[1]]$model_base
Call:
coxph(formula = as.formula(paste("Surv(df_train$time, df_train$event) ~",
    paste(predict.factors, collapse = "+"))), data = df_train,
    x = TRUE)

            coef exp(coef)  se(coef)      z        p
age     0.045127  1.046160  0.008373  5.389 7.07e-08
gender -0.235510  0.790168  0.177216 -1.329  0.18387
hr      0.011451  1.011516  0.003917  2.923  0.00347
sysbp   0.003548  1.003555  0.003670  0.967  0.33358
diasbp -0.013238  0.986849  0.006395 -2.070  0.03845
bmi    -0.047169  0.953926  0.019425 -2.428  0.01517
cvd    -0.025967  0.974367  0.219483 -0.118  0.90582
afb     0.046888  1.048005  0.224161  0.209  0.83431
sho     0.942797  2.567151  0.366438  2.573  0.01009
chf     0.518465  1.679448  0.189360  2.738  0.00618
av3     0.550735  1.734527  0.490364  1.123  0.26139
miord   0.031701  1.032208  0.189273  0.167  0.86699
mitype -0.028336  0.972062  0.230706 -0.123  0.90225
los    -0.023947  0.976338  0.024218 -0.989  0.32277
```

```
y1997  -0.205430  0.814297  0.245965 -0.835  0.40360
y1999  -0.185491  0.830697  0.230437 -0.805  0.42085

Likelihood ratio test=135.4  on 16 df, p=< 2.2e-16
n= 333, number of events= 148


$model_list[[2]]
$model_list[[2]]$beststats
   mtry nodesize nodedepth time    AUCROC        BS
V1    3       30        50  2.9 0.8485461 0.1693326
   BS_scaled   C_score Calib_alpha Calib_slope
V1 0.3298849 0.7695282  0.03475621     1.61813

$model_list[[2]]$allstats
    mtry nodesize nodedepth time    AUCROC        BS
V1     4       15        50  2.9 0.8470288 0.1645384
V2     4       20        50  2.9 0.8463550 0.1658128
V3     4       25        50  2.9 0.8463719 0.1663581
V4     4       30        50  2.9 0.8476221 0.1672425
V11    3       30        50  2.9 0.8485461 0.1693326
V21    5       30        50  2.9 0.8468629 0.1658770
V31    7       30        50  2.9 0.8442001 0.1647465
V41   10       30        50  2.9 0.8432322 0.1638070
V5    15       30        50  2.9 0.8365888 0.1630288
    BS_scaled   C_score Calib_alpha Calib_slope
V1  0.3488577 0.7701725  0.02977282   1.2259684
V2  0.3438143 0.7692060  0.03035507   1.2738710
V3  0.3416565 0.7705825  0.03308607   1.3450685
V4  0.3381564 0.7715198  0.03512416   1.4405591
V11 0.3298849 0.7695282  0.03475621   1.6181298
V21 0.3435602 0.7697332  0.03223757   1.3411137
V31 0.3480343 0.7693524  0.03321698   1.1814584
V41 0.3517520 0.7656621  0.03162826   1.0670551
V5  0.3548316 0.7616202  0.03361119   0.9803958

$model_list[[2]]$model
                         Sample size: 334
                    Number of deaths: 146
                     Number of trees: 500
           Forest terminal node size: 30
       Average no. of terminal nodes: 8.546
No. of variables tried at each split: 3
              Total no. of variables: 17
       Resampling used to grow trees: swor
    Resample size used to grow trees: 211
                            Analysis: RSF
                              Family: surv
                    Splitting rule: logrank *random*
       Number of random split points: 50
                          (OOB) CRPS: 0.15583045
   (OOB) Requested performance error: 0.22954033


$model_list[[2]]$vimp10
  cox_predict           age           chf           sho
```

```
 8.027918e-02  3.106505e-02  1.965666e-02  9.984943e-03
          bmi            hr           afb         sysbp
 7.124829e-03  4.445464e-03  1.601209e-03  1.182673e-03
          los        diasbp         y1997        gender
 1.154096e-03  1.151968e-03  7.810680e-04  5.613834e-04
        y1999           cvd        mitype
 5.823523e-05 -1.111680e-04 -1.658726e-04

$model_list[[2]]$model_base
Call:
coxph(formula = as.formula(paste("Surv(df_train$time, df_train$event) ~",
    paste(predict.factors, collapse = "+"))), data = df_train,
    x = TRUE)

             coef  exp(coef)   se(coef)      z        p
age     0.0427255  1.0436514  0.0082711  5.166 2.40e-07
gender -0.4636530  0.6289817  0.1845558 -2.512  0.01200
hr      0.0105162  1.0105717  0.0040092  2.623  0.00871
sysbp  -0.0016290  0.9983723  0.0036701 -0.444  0.65715
diasbp -0.0124902  0.9875874  0.0060900 -2.051  0.04027
bmi    -0.0628271  0.9391058  0.0221249 -2.840  0.00452
cvd     0.0550500  1.0565934  0.2185633  0.252  0.80114
afb    -0.0564619  0.9451025  0.2190806 -0.258  0.79662
sho     1.5720199  4.8163671  0.3514348  4.473 7.71e-06
chf     0.8723982  2.3926420  0.2024632  4.309 1.64e-05
av3    -0.0124767  0.9876008  0.6267103 -0.020  0.98412
miord  -0.0127836  0.9872978  0.1875898 -0.068  0.94567
mitype -0.2589697  0.7718464  0.2389325 -1.084  0.27843
los     0.0008633  1.0008637  0.0176832  0.049  0.96106
y1997  -0.6730080  0.5101717  0.2518453 -2.672  0.00753
y1999  -0.3506881  0.7042033  0.2262112 -1.550  0.12108

Likelihood ratio test=172.6  on 16 df, p=< 2.2e-16
n= 334, number of events= 146


$model_list[[3]]
$model_list[[3]]$beststats
   mtry nodesize nodedepth time   AUCROC        BS
V1    3       15        50  2.2 0.838926 0.1551005
   BS_scaled   C_score Calib_alpha Calib_slope
V1 0.3236312 0.7741169   0.0307184    1.328401

$model_list[[3]]$allstats
    mtry nodesize nodedepth time    AUCROC        BS
V1     4       15        50  2.2 0.8378846 0.1534363
V2     4       20        50  2.2 0.8357532 0.1542568
V3     4       25        50  2.2 0.8357496 0.1553463
V4     4       30        50  2.2 0.8356515 0.1560454
V11    3       15        50  2.2 0.8389260 0.1551005
V21    5       15        50  2.2 0.8362526 0.1541165
V31    7       15        50  2.2 0.8346919 0.1552187
V41   10       15        50  2.2 0.8308972 0.1577378
V5    15       15        50  2.2 0.8215753 0.1628212
    BS_scaled   C_score Calib_alpha Calib_slope
V1  0.3308887 0.7733148  0.03260733   1.2015905
```

```
V2  0.3273104 0.7714638  0.03589434   1.2563099
V3  0.3225595 0.7712479  0.03674847   1.3048902
V4  0.3195106 0.7707543  0.03668933   1.3801161
V11 0.3236312 0.7741169  0.03071840   1.3284013
V21 0.3279225 0.7727904  0.02938729   1.1057494
V31 0.3231157 0.7749807  0.02703298   1.0236751
V41 0.3121303 0.7722659  0.02596653   0.9024273
V5  0.2899623 0.7666821  0.02500830   0.8013229

$model_list[[3]]$model
                         Sample size: 333
                    Number of deaths: 136
                     Number of trees: 500
          Forest terminal node size: 15
       Average no. of terminal nodes: 16.426
No. of variables tried at each split: 3
              Total no. of variables: 17
       Resampling used to grow trees: swor
    Resample size used to grow trees: 210
                            Analysis: RSF
                              Family: surv
                    Splitting rule: logrank *random*
       Number of random split points: 50
                          (OOB) CRPS: 0.15684567
   (OOB) Requested performance error: 0.22876109


$model_list[[3]]$vimp10
 cox_predict          age          chf          bmi
0.0610714613 0.0404930999 0.0177301208 0.0092150952
          hr          los       gender       diasbp
0.0052743281 0.0043323313 0.0018941710 0.0014710194
         sho        y1997        sysbp        miord
0.0013158923 0.0005718562 0.0003803418 0.0003494961
         cvd       mitype          av3
0.0003476008 0.0002391834 0.0001877771

$model_list[[3]]$model_base
Call:
coxph(formula = as.formula(paste("Surv(df_train$time, df_train$event) ~",
    paste(predict.factors, collapse = "+"))), data = df_train,
    x = TRUE)

             coef  exp(coef)   se(coef)      z        p
age     0.0537301  1.0551998  0.0095672  5.616 1.95e-08
gender -0.2346093  0.7908797  0.1822319 -1.287 0.197947
hr      0.0112430  1.0113064  0.0038730  2.903 0.003697
sysbp   0.0012513  1.0012521  0.0036191  0.346 0.729524
diasbp -0.0129867  0.9870973  0.0059917 -2.167 0.030202
bmi    -0.0371948  0.9634884  0.0207787 -1.790 0.073447
cvd    -0.1519318  0.8590468  0.2385936 -0.637 0.524268
afb     0.2506873  1.2849082  0.2343887  1.070 0.284828
sho     1.4600417  4.3061390  0.3960059  3.687 0.000227
chf     0.8340849  2.3027060  0.1887787  4.418 9.95e-06
av3    -0.0467745  0.9543026  0.5596070 -0.084 0.933387
miord   0.1944781  1.2146769  0.1852333  1.050 0.293760
```

```
mitype -0.4134881  0.6613394  0.2645408 -1.563 0.118043
los     0.0009399  1.0009404  0.0200538  0.047 0.962617
y1997  -0.4720247  0.6237381  0.2512653 -1.879 0.060300
y1999  -0.4348502  0.6473617  0.2340114 -1.858 0.063134

Likelihood ratio test=166.9  on 16 df, p=< 2.2e-16
n= 333, number of events= 136



$time
Time difference of 6.509842 secs
```