Simulated examples for the survival ensemble methods

Select data type

Hnscc

Time point for event prediction:

4

Random seed for calibration and validation

43

K_Outer loop CV (for validation)

5

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	0.71	0.1095	-0.0335	0.7097	0.7545	-0.0009	4
train	0.9461	0.0687	0.3754	0.9249	2.9745	-0.0003	4

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	0.5826	0.1167	-0.4341	0.5843	0.1935	-0.0671	4
test.2	0.7381	0.1304	0.1161	0.7541	0.8297	0.0558	4
test.3	0.6896	0.0895	-0.0633	0.6751	0.6161	-0.0396	4
test.4	0.6535	0.1393	0.0014	0.6533	0.3864	0.0579	4
test.5	0.8861	0.0715	0.2125	0.8816	1.7468	-0.0117	4

Showing 1 to 5 of 5 entries Previous 1 Next

```
  |
  |                                                      |   0%
  |
  |=======                                               |  14%
  |
  |=============                                         |  29%
  |
  |====================                                  |  43%
  |
  |===========================                           |  57%
  |
  |==================================                    |  71%
  |
  |==========================================            |  86%
  |
  |======================================================| 100%
$test
  T    AUCROC         BS    BS_scaled   C_score
1 4 0.5825863 0.11670088 -0.434076656 0.5843195
2 4 0.7381022 0.13043801  0.116129772 0.7541112
3 4 0.6895977 0.08951871 -0.063269916 0.6750742
4 4 0.6534935 0.13927717  0.001445078 0.6533115
5 4 0.8860530 0.07146151  0.212467813 0.8815612
  Calib_slope Calib_alpha test cv_n
1   0.1935432 -0.06710218    1    1
2   0.8297293  0.05584705    1    2
3   0.6161315 -0.03961375    1    3
4   0.3864348  0.05787541    1    4
5   1.7467566 -0.01166590    1    5

$train
  T    AUCROC         BS BS_scaled   C_score Calib_slope
1 4 0.9805320 0.05480619 0.5317606 0.9564613    3.516517
2 4 0.9607155 0.06436830 0.3563793 0.9373364    3.559308
3 4 0.9316559 0.07510530 0.3584907 0.9134112    2.397352
4 4 0.9326939 0.06921912 0.3262703 0.9138563    2.572772
5 4 0.9247811 0.08000463 0.3042103 0.9035897    2.826558
    Calib_alpha test cv_n
1 -2.273377e-03    0    1
2  4.883290e-04    0    2
3 -6.746421e-04    0    3
4  7.556387e-05    0    4
5  6.343831e-04    0    5

$testaverage
            T        AUCROC            BS     BS_scaled
 4.0000000000  0.7099665478  0.1094792570 -0.0334607819
      C_score   Calib_slope   Calib_alpha          test
 0.7096755352  0.7545190856 -0.0009318736  1.0000000000

$trainaverage
            T        AUCROC            BS     BS_scaled
 4.0000000000  0.9460757033  0.0687007052  0.3754222469
      C_score   Calib_slope   Calib_alpha          test
```

```
 0.9249309878  2.9745015660 -0.0003499487  0.0000000000

$model_list
$model_list[[1]]
$model_list[[1]]$beststats
   mtry nodesize nodedepth time    AUCROC         BS
V2   21       15        50  3.3 0.8203674 0.08827607
   BS_scaled   C_score Calib_alpha Calib_slope
V2 0.1101757 0.7509637 -0.01319374   0.7533519

$model_list[[1]]$allstats
    mtry nodesize nodedepth time    AUCROC         BS
V1    10       15        50  3.3 0.8131007 0.08769535
V2    10       20        50  3.3 0.8126556 0.08694113
V3    10       25        50  3.3 0.8118626 0.08643769
V4    10       30        50  3.3 0.8126215 0.08601591
V5    10       35        50  3.3 0.8102518 0.08583867
V11   11       15        50  3.3 0.8149546 0.08647083
V21   21       15        50  3.3 0.8203674 0.08827607
V31   36       15        50  3.3 0.8104541 0.09143210
V41   54       15        50  3.3 0.8090970 0.09124450
V51   10       15        50  3.3 0.8131007 0.08769535
     BS_scaled   C_score  Calib_alpha Calib_slope
V1  0.11602941 0.7424961 -0.009544008   0.8039529
V2  0.12363194 0.7435703 -0.008655802   0.8378539
V3  0.12870658 0.7422749 -0.007062857   0.8629763
V4  0.13295815 0.7416114 -0.006063651   0.8934772
V5  0.13474471 0.7412006 -0.005920419   0.9345742
V11 0.12837253 0.7418009 -0.010727954   0.8168714
V21 0.11017572 0.7509637 -0.013193741   0.7533519
V31 0.07836287 0.7413902 -0.014998827   0.6670799
V41 0.08025385 0.7408215 -0.017103684   0.6393121
V51 0.11602941 0.7424961 -0.009544008   0.8039529

$model_list[[1]]$model
                         Sample size: 361
                    Number of deaths: 48
                     Number of trees: 500
           Forest terminal node size: 15
       Average no. of terminal nodes: 15.172
No. of variables tried at each split: 21
              Total no. of variables: 107
       Resampling used to grow trees: swor
    Resample size used to grow trees: 228
                            Analysis: RSF
                              Family: surv
                     Splitting rule: logrank *random*
       Number of random split points: 50
                          (OOB) CRPS: 0.06203216
   (OOB) Requested performance error: 0.25475453



$model_list[[2]]
$model_list[[2]]$beststats
   mtry nodesize nodedepth time    AUCROC         BS
```

```
V5   10       25        50  2.9 0.7570005 0.08202887
    BS_scaled   C_score  Calib_alpha Calib_slope
V5 0.04512564 0.7032827 -0.006857878   0.7356402

$model_list[[2]]$allstats
    mtry nodesize nodedepth time    AUCROC         BS
V1    10       15        50  2.9 0.7406581 0.08513219
V2    10       20        50  2.9 0.7559812 0.08261163
V3    10       25        50  2.9 0.7570005 0.08202887
V4    10       30        50  2.9 0.7508127 0.08237271
V5    10       35        50  2.9 0.7479396 0.08214145
V11   11       25        50  2.9 0.7511480 0.08363471
V21   21       25        50  2.9 0.7435998 0.08545408
V31   36       25        50  2.9 0.7457475 0.08692553
V41   54       25        50  2.9 0.7416239 0.08697794
V51   10       25        50  2.9 0.7570005 0.08202887
       BS_scaled   C_score  Calib_alpha Calib_slope
V1   0.009000752 0.6911688 -0.009886011   0.6032343
V2   0.038341832 0.7049278 -0.008646927   0.7011229
V3   0.045125638 0.7032827 -0.006857878   0.7356402
V4   0.041123118 0.6976744 -0.005719604   0.7524031
V5   0.043815058 0.6952815 -0.005204440   0.7884184
V11  0.026432507 0.6999177 -0.005371366   0.6939480
V21  0.005253791 0.6904584 -0.008299573   0.6168618
V31 -0.011874989 0.6955433 -0.011180077   0.5536702
V41 -0.012485102 0.6899349 -0.010101583   0.5099371
V51  0.045125638 0.7032827 -0.006857878   0.7356402

$model_list[[2]]$model
                         Sample size: 361
                    Number of deaths: 40
                     Number of trees: 500
           Forest terminal node size: 25
       Average no. of terminal nodes: 10.356
No. of variables tried at each split: 10
              Total no. of variables: 107
       Resampling used to grow trees: swor
    Resample size used to grow trees: 228
                            Analysis: RSF
                              Family: surv
                     Splitting rule: logrank *random*
       Number of random split points: 50
                           (OOB) CRPS: 0.05472273
   (OOB) Requested performance error: 0.29785405



$model_list[[3]]
$model_list[[3]]$beststats
   mtry nodesize nodedepth time    AUCROC         BS
V2   21       35        50  2.9 0.7496691 0.09539907
    BS_scaled   C_score  Calib_alpha Calib_slope
V2 0.03706826 0.7044809 -0.009536267   0.6284023

$model_list[[3]]$allstats
    mtry nodesize nodedepth time    AUCROC         BS
```

```
V1    10       15        50  2.9 0.7397643 0.09631209
V2    10       20        50  2.9 0.7411764 0.09570050
V3    10       25        50  2.9 0.7467761 0.09425049
V4    10       30        50  2.9 0.7484842 0.09396004
V5    10       35        50  2.9 0.7494832 0.09372072
V11   11       35        50  2.9 0.7464931 0.09435143
V21   21       35        50  2.9 0.7496691 0.09539907
V31   36       35        50  2.9 0.7398922 0.09826334
V41   54       35        50  2.9 0.7365362 0.09971820
V51   10       35        50  2.9 0.7494832 0.09372072
       BS_scaled   C_score  Calib_alpha Calib_slope
V1   0.027852434 0.7003471 -0.011425658   0.5835916
V2   0.034025694 0.7000000 -0.009168628   0.6097645
V3   0.048661687 0.7042916 -0.008079083   0.6667808
V4   0.051593369 0.7046071 -0.007673451   0.6833091
V5   0.054009017 0.7054591 -0.007200453   0.7097917
V11  0.047642823 0.7036920 -0.007992866   0.6878306
V21  0.037068258 0.7044809 -0.009536267   0.6284023
V31  0.008157060 0.6981067 -0.010843819   0.5447102
V41 -0.006527855 0.6954244 -0.015223275   0.4757649
V51  0.054009017 0.7054591 -0.007200453   0.7097917

$model_list[[3]]$model
                         Sample size: 361
                    Number of deaths: 48
                     Number of trees: 500
           Forest terminal node size: 35
       Average no. of terminal nodes: 11.588
No. of variables tried at each split: 21
              Total no. of variables: 107
       Resampling used to grow trees: swor
    Resample size used to grow trees: 228
                            Analysis: RSF
                              Family: surv
                    Splitting rule: logrank *random*
       Number of random split points: 50
                          (OOB) CRPS: 0.06385849
   (OOB) Requested performance error: 0.28926611



$model_list[[4]]
$model_list[[4]]$beststats
   mtry nodesize nodedepth time    AUCROC         BS
V1   11       35        50  3.1 0.7765149 0.08176475
    BS_scaled   C_score  Calib_alpha Calib_slope
V1 0.07532043 0.7325513 -0.002115992   0.8186295

$model_list[[4]]$allstats
    mtry nodesize nodedepth time    AUCROC         BS
V1    10       15        50  3.1 0.7572123 0.08343519
V2    10       20        50  3.1 0.7619155 0.08271453
V3    10       25        50  3.1 0.7663508 0.08245395
V4    10       30        50  3.1 0.7740229 0.08160910
V5    10       35        50  3.1 0.7757574 0.08125640
V11   11       35        50  3.1 0.7765149 0.08176475
```

```
V21   21       35        50  3.1 0.7744627 0.08255439
V31   36       35        50  3.1 0.7653990 0.08389747
V41   54       35        50  3.1 0.7650306 0.08529124
V51   10       35        50  3.1 0.7757574 0.08125640
     BS_scaled   C_score  Calib_alpha Calib_slope
V1  0.05642946 0.7198680 -0.007203848   0.6582768
V2  0.06457937 0.7231672 -0.005011774   0.7257167
V3  0.06752633 0.7280059 -0.003904156   0.7646603
V4  0.07708073 0.7319648 -0.002747276   0.8259253
V5  0.08106940 0.7337243 -0.002098028   0.8684682
V11 0.07532043 0.7325513 -0.002115992   0.8186295
V21 0.06639045 0.7362170 -0.003674148   0.7286030
V31 0.05120149 0.7299120 -0.005651153   0.6366882
V41 0.03543929 0.7331378 -0.008872823   0.6029357
V51 0.08106940 0.7337243 -0.002098028   0.8684682

$model_list[[4]]$model
                         Sample size: 360
                    Number of deaths: 41
                     Number of trees: 500
          Forest terminal node size: 35
       Average no. of terminal nodes: 9.474
No. of variables tried at each split: 11
              Total no. of variables: 107
       Resampling used to grow trees: swor
    Resample size used to grow trees: 228
                            Analysis: RSF
                              Family: surv
                     Splitting rule: logrank *random*
       Number of random split points: 50
                          (OOB) CRPS: 0.05443377
   (OOB) Requested performance error: 0.27259732



$model_list[[5]]
$model_list[[5]]$beststats
   mtry nodesize nodedepth time    AUCROC         BS
V5   10       30        50  2.9 0.7193498 0.09494391
    BS_scaled   C_score  Calib_alpha Calib_slope
V5 0.02019391 0.6727564 -0.003485748   0.6662571

$model_list[[5]]$allstats
    mtry nodesize nodedepth time    AUCROC         BS
V1    10       15        50  2.9 0.7091161 0.09639563
V2    10       20        50  2.9 0.7162945 0.09594669
V3    10       25        50  2.9 0.7175232 0.09569405
V4    10       30        50  2.9 0.7193498 0.09494391
V5    10       35        50  2.9 0.7159702 0.09504237
V11   11       30        50  2.9 0.7100608 0.09653943
V21   21       30        50  2.9 0.7075015 0.09910962
V31   36       30        50  2.9 0.7047155 0.09951178
V41   54       30        50  2.9 0.6961444 0.10294911
V51   10       30        50  2.9 0.7193498 0.09494391
       BS_scaled   C_score  Calib_alpha Calib_slope
V1   0.005212348 0.6635897 -0.007947007   0.5555321
```

```
V2   0.009845394 0.6682051 -0.006171300   0.6038977
V3   0.012452547 0.6698718 -0.004539917   0.6276066
V4   0.020193907 0.6727564 -0.003485748   0.6662571
V5   0.019177808 0.6687821 -0.003123228   0.6777664
V11  0.003728345 0.6665385 -0.006383028   0.6109346
V21 -0.022795635 0.6651282 -0.008189335   0.5254106
V31 -0.026945802 0.6634615 -0.009160834   0.4920373
V41 -0.062418514 0.6575962 -0.012686798   0.3869847
V51  0.020193907 0.6727564 -0.003485748   0.6662571

$model_list[[5]]$model
                         Sample size: 361
                    Number of deaths: 47
                     Number of trees: 500
          Forest terminal node size: 30
       Average no. of terminal nodes: 11.018
No. of variables tried at each split: 10
              Total no. of variables: 107
       Resampling used to grow trees: swor
    Resample size used to grow trees: 228
                            Analysis: RSF
                              Family: surv
                    Splitting rule: logrank *random*
       Number of random split points: 50
                          (OOB) CRPS: 0.06494129
   (OOB) Requested performance error: 0.32360105




$time
Time difference of 11.05891 secs
```