CI-project-chemical-cart

2022-12-08

```r
library(propensityml)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
library(sas7bdat)
```

Data

```r
df<-read.table('/Users/kyungseonlee/Desktop/causal/poisox.txt')
colnames(df)=c('age','sex','prior','poisox','subseq','mortal')
df%>% head()
```

```
##   age sex prior poisox subseq mortal
## 1  31   0   126      1    140      0
## 2  35   1   115      1    124      1
## 3  27   1    92      0    101      0
## 4  66   0   118      0    112      1
## 5  54   1    89      1     90      0
## 6  35   1    93      1    104      0
```

```r
df3=df
df3$blood_diff=df3$subseq-df3$prior
df3=df3[c(1,2,4,6,7)]
df3 %>% head()
```

```
##   age sex poisox mortal blood_diff
## 1  31   0      1      0         14
## 2  35   1      1      1          9
## 3  27   1      0      0          9
## 4  66   0      0      1         -6
## 5  54   1      1      0          1
## 6  35   1      1      0         11
```

```r
library(rpart)

#control = rpart.control(minbucket = 2)
df1_c=df3
df1_c %>% head(5)
```

```
##   age sex poisox mortal blood_diff
## 1  31   0      1      0         14
## 2  35   1      1      1          9
## 3  27   1      0      0          9
## 4  66   0      0      1         -6
## 5  54   1      1      0          1
```

```r
cart2=rpart(poisox~ age+sex+blood_diff, data=df1_c ,method='poisson')
summary(cart2)
```

```
## Call:
## rpart(formula = poisox ~ age + sex + blood_diff, data = df1_c,
##     method = "poisson")
##   n= 5000
##
##           CP nsplit  rel error     xerror       xstd
## 1 0.46629674      0 1.00000000 1.00037840 0.01808746
## 2 0.07551949      1 0.53370326 0.53417044 0.01884740
## 3 0.06519506      3 0.38266427 0.38927042 0.01635065
## 4 0.05989831      5 0.25227415 0.25886521 0.01528843
## 5 0.01873951      7 0.13247753 0.14201198 0.01345788
## 6 0.01098757      8 0.11373802 0.12409004 0.01133002
## 7 0.01000000     11 0.07935346 0.09866446 0.01038412
##
## Variable importance
## blood_diff        age
##         87         13
##
## Node number 1: 5000 observations,    complexity param=0.4662967
##   events=3604,  estimated rate=0.7208 , mean deviance=0.4719706
##   left son=2 (779 obs) right son=3 (4221 obs)
##   Primary splits:
##       blood_diff < -2.5 to the left,  improve=1100.46700, (0 missing)
##       sex        < 0.5  to the left,  improve=  60.09432, (0 missing)
##       age        < 25.5 to the left,  improve=  29.14106, (0 missing)
##
## Node number 2: 779 observations,    complexity param=0.01873951
##   events=12,  estimated rate=0.01665839 , mean deviance=0.1286649
##   left son=4 (704 obs) right son=5 (75 obs)
##   Primary splits:
##       blood_diff < -3.5 to the left,  improve=44.80991, (0 missing)
##       sex        < 0.5  to the left,  improve=13.17852, (0 missing)
##       age        < 60.5 to the right, improve=10.49810, (0 missing)
##   Surrogate splits:
##       age < 38.5 to the right, agree=0.919, adj=0.16, (0 split)
##
## Node number 3: 4221 observations,    complexity param=0.07551949
##   events=3592,  estimated rate=0.8509404 , mean deviance=0.2746343
##   left son=6 (1673 obs) right son=7 (2548 obs)
##   Primary splits:
##       age        < 36.5 to the left,  improve=127.58650, (0 missing)
##       blood_diff < 8.5  to the right, improve= 23.16781, (0 missing)
##       sex        < 0.5  to the left,  improve= 15.24260, (0 missing)
##   Surrogate splits:
##       blood_diff < 8.5  to the right, agree=0.924, adj=0.808, (0 split)
##
## Node number 4: 704 observations
##   events=1,  estimated rate=0.002835322 , mean deviance=0.01949337
##
## Node number 5: 75 observations
##   events=11,  estimated rate=0.1570941 , mean deviance=0.5637885
##
## Node number 6: 1673 observations,    complexity param=0.07551949
##   events=1098,  estimated rate=0.6563595 , mean deviance=0.5527778
##   left son=12 (180 obs) right son=13 (1493 obs)
```

```
##    Primary splits:
##        blood_diff < 6.5  to the left,  improve=229.20630, (0 missing)
##        sex        < 0.5  to the left,  improve= 32.12424, (0 missing)
##        age        < 25.5 to the left,  improve= 16.30528, (0 missing)
##
## Node number 7: 2548 observations,    complexity param=0.01098757
##   events=2494,  estimated rate=0.9786665 , mean deviance=0.04193385
##   left son=14 (98 obs) right son=15 (2450 obs)
##    Primary splits:
##        blood_diff < -1.5 to the left,  improve=21.3570900, (0 missing)
##        age        < 68.5 to the right, improve= 4.8083430, (0 missing)
##        sex        < 0.5  to the left,  improve= 0.1394467, (0 missing)
##
## Node number 12: 180 observations
##   events=2,  estimated rate=0.01653919 , mean deviance=0.1020122
##
## Node number 13: 1493 observations,    complexity param=0.06519506
##   events=1096,  estimated rate=0.7340801 , mean deviance=0.4538458
##   left son=26 (963 obs) right son=27 (530 obs)
##    Primary splits:
##        age        < 29.5 to the left,  improve=62.77969, (0 missing)
##        blood_diff < 23.5 to the left,  improve=25.76237, (0 missing)
##        sex        < 0.5  to the left,  improve=15.21969, (0 missing)
##    Surrogate splits:
##        blood_diff < 15.5 to the right, agree=0.821, adj=0.496, (0 split)
##
## Node number 14: 98 observations,    complexity param=0.01098757
##   events=55,  estimated rate=0.563452 , mean deviance=0.6483738
##   left son=28 (25 obs) right son=29 (73 obs)
##    Primary splits:
##        age < 63   to the right, improve=32.395880, (0 missing)
##        sex < 0.5  to the left,  improve= 1.274737, (0 missing)
##
## Node number 15: 2450 observations
##   events=2439,  estimated rate=0.9953547 , mean deviance=0.008959428
##
## Node number 26: 963 observations,    complexity param=0.06519506
##   events=579,  estimated rate=0.6014181 , mean deviance=0.6117691
##   left son=52 (198 obs) right son=53 (765 obs)
##    Primary splits:
##        blood_diff < 15.5 to the left,  improve=245.286400, (0 missing)
##        sex        < 0.5  to the left,  improve= 21.239770, (0 missing)
##        age        < 24.5 to the left,  improve=  7.422354, (0 missing)
##
## Node number 27: 530 observations
##   events=517,  estimated rate=0.9748068 , mean deviance=0.04845044
##
## Node number 28: 25 observations
##   events=0,  estimated rate=0.03789695 , mean deviance=0.0757939
##
## Node number 29: 73 observations,    complexity param=0.01098757
##   events=55,  estimated rate=0.7528162 , mean deviance=0.4266291
##   left son=58 (18 obs) right son=59 (55 obs)
##    Primary splits:
##        age < 43.5 to the left,  improve=31.1438900, (0 missing)
##        sex < 0.5  to the left,  improve= 0.6024904, (0 missing)
```

```
##
## Node number 52: 198 observations
##   events=2,  estimated rate=0.01504609 , mean deviance=0.09467071
##
## Node number 53: 765 observations,    complexity param=0.05989831
##   events=577,  estimated rate=0.7541878 , mean deviance=0.4254467
##   left son=106 (419 obs) right son=107 (346 obs)
##   Primary splits:
##       age        < 24.5 to the left,  improve=45.719020, (0 missing)
##       blood_diff < 22.5 to the left,  improve=29.525310, (0 missing)
##       sex        < 0.5  to the left,  improve= 7.452957, (0 missing)
##   Surrogate splits:
##       blood_diff < 25.5 to the right, agree=0.722, adj=0.384, (0 split)
##       sex        < 0.5  to the left,  agree=0.554, adj=0.014, (0 split)
##
## Node number 58: 18 observations
##   events=0,  estimated rate=0.05158003 , mean deviance=0.1031601
##
## Node number 59: 55 observations
##   events=55,  estimated rate=0.9931306 , mean deviance=4.740592e-05
##
## Node number 106: 419 observations,    complexity param=0.05989831
##   events=235,  estimated rate=0.561387 , mean deviance=0.6486739
##   left son=212 (167 obs) right son=213 (252 obs)
##   Primary splits:
##       blood_diff < 22.5 to the left,  improve=238.967700, (0 missing)
##       sex        < 0.5  to the left,  improve=  8.540399, (0 missing)
##       age        < 23.5 to the left,  improve=  3.622450, (0 missing)
##
## Node number 107: 346 observations
##   events=342,  estimated rate=0.9873704 , mean deviance=0.02298838
##
## Node number 212: 167 observations
##   events=0,  estimated rate=0.005938688 , mean deviance=0.01187738
##
## Node number 213: 252 observations
##   events=235,  estimated rate=0.9313804 , mean deviance=0.1302652
```

```
cart2$control
```

```
## $minsplit
## [1] 20
##
## $minbucket
## [1] 7
##
## $cp
## [1] 0.01
##
## $maxcompete
## [1] 4
##
## $maxsurrogate
## [1] 5
##
## $usesurrogate
## [1] 2
##
## $surrogatestyle
## [1] 0
##
## $maxdepth
## [1] 30
##
## $xval
## [1] 10
```

```
cart_ps2=predict(cart2)
cart_ps2 %>% head()
```

```
##                1           2           3           4           5           6
## 0.974806801 0.974806801 0.015046090 0.002835322 0.995354734 0.974806801
```

```
df1_c[,"ps2"]=cart_ps2

df1_c %>% head()
```

```
##    age sex poisox mortal blood_diff         ps2
## 1  31   0      1      0         14 0.974806801
## 2  35   1      1      1          9 0.974806801
## 3  27   1      0      0          9 0.015046090
## 4  66   0      0      1         -6 0.002835322
## 5  54   1      1      0          1 0.995354734
## 6  35   1      1      0         11 0.974806801
```
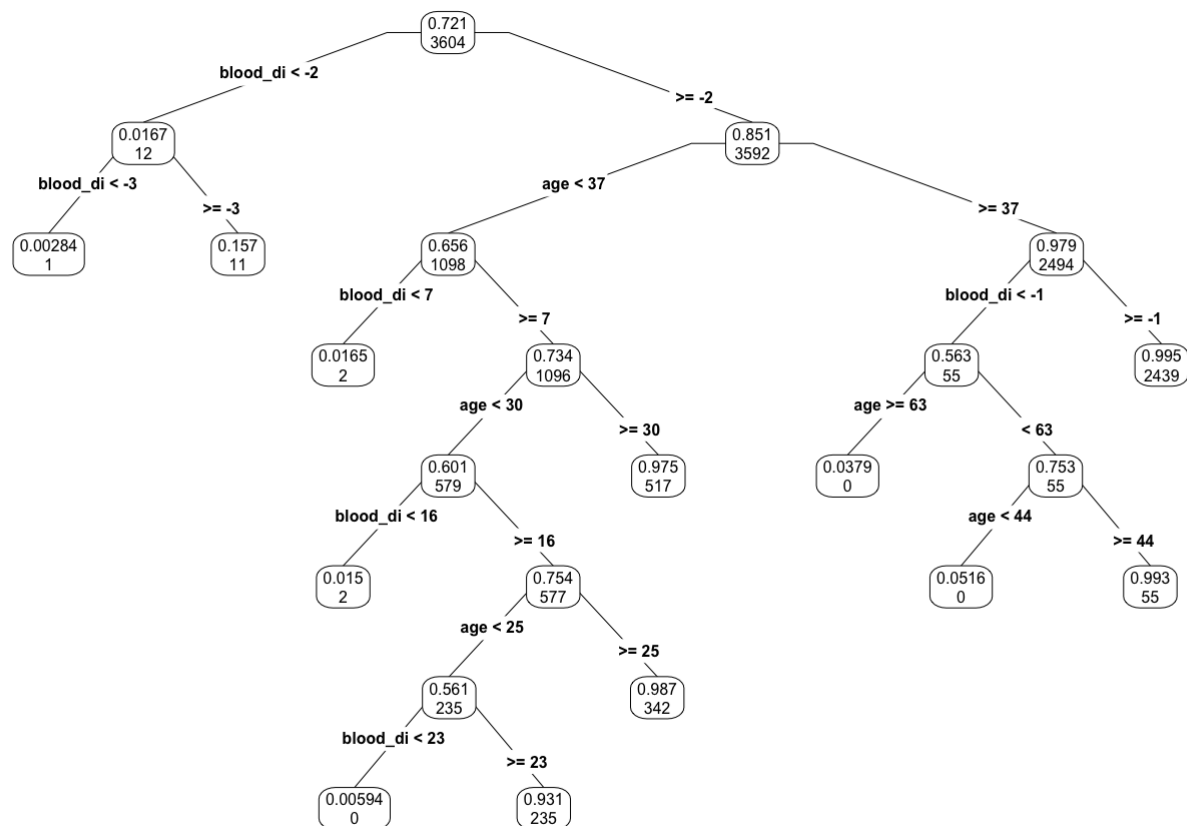
```
sum(df1_c$ps2==0|df1_c$ps2==1)
```

```
## [1] 0
```

```
#install.packages("rpart.plot")

library(rpart.plot)

prp(cart2, type=4, extra=2, digits=3)
```



```
zi=df1_c$poisox
yi=df1_c$mortal
e=df1_c$ps2


df1_c["ipw_wt"]= zi/e-(1-zi)/(1-e)
df1_c %>% head()
```

```
##    age sex poisox mortal blood_diff         ps2     ipw_wt
## 1  31   0      1      0         14 0.974806801   1.025844
## 2  35   1      1      1          9 0.974806801   1.025844
## 3  27   1      0      0          9 0.015046090  -1.015276
## 4  66   0      0      1         -6 0.002835322  -1.002843
## 5  54   1      1      0          1 0.995354734   1.004667
## 6  35   1      1      0         11 0.974806801   1.025844
```

```
ATE_ipw_cart=mean(zi*yi/e)-mean((1-zi)*yi/(1-e))
ATE_ipw_cart
```

```
## [1] -0.05493321
```

```
ATE_sipw_cart=sum(zi*yi/e)/sum(zi/e)-sum((1-zi)*yi/(1-e))/sum((1-zi)/(1-e))
ATE_sipw_cart
```

```
## [1] -0.004002915
```

```
cov_balance_c=data.frame(rep(0),row.names = "CART")

df3%>% head()
```

```
##   age sex poisox mortal blood_diff
## 1  31   0      1      0         14
## 2  35   1      1      1          9
## 3  27   1      0      0          9
## 4  66   0      0      1         -6
## 5  54   1      1      0          1
## 6  35   1      1      0         11
```

```
for(i in colnames(df3)){
  if(i!="poisox" & i!="mortal"){
#    print(df1[i])
    t_weighted_mean=mean((df3[i]*df1_c$ipw_wt)[df3$poisox==1,])
    c_weighted_mean=mean((df3[i]*df1_c$ipw_wt)[df3$poisox==0,])
    weighted_mean_diff=abs(t_weighted_mean-c_weighted_mean)
    asam=weighted_mean_diff/sd((df3[i]*df1_c$ipw_wt)[df3$poisox==1,])
    cov_balance_c[i]=asam
  }
}
cov_balance_c=cov_balance_c[,-1]
cov_balance_c
```

```
##          age       sex blood_diff
## CART 0.61485 0.3193976  0.4773585
```

```
cov_balance_c$ASAM=apply(cov_balance_c,1,mean)
cov_balance_c["ASAM"]
```

```
##            ASAM
## CART 0.4705354
```