Data Analysis Homework 2

Jimmy Hickey

9/25/2020

1

```
library(DynTxRegime)
```

```
## Loading required package: modelObj
```

```
data <- read.csv(file ="cholesterol.dat.txt", header = TRUE, sep = ",")
data$A= data$trt

y = data$chol0 - data$chol6

lm = buildModelObj(model = ~A + exercise + wt + smoke + trig0 + age + gender +
                     A:exercise + A:wt + A:smoke + A:trig0 + A:age + A:gender,
                   solver.method = "lm",
                   predict.method = "predict.lm",
                   predict.args = list("type"="response"))
# adj R^2 = 0.889

summary(fit(lm, data, y))
```

```
##
## Call:
## lm(formula = YinternalY ~ A + exercise + wt + smoke + trig0 +
##     age + gender + A:exercise + A:wt + A:smoke + A:trig0 + A:age +
##     A:gender, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -73.049  -8.053   0.449   8.693  35.895
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.246e+01  5.560e+00   5.839 7.14e-09 ***
## A           -2.615e+02  8.215e+00 -31.825  < 2e-16 ***
## exercise     2.058e+01  1.792e+00  11.489  < 2e-16 ***
## wt          -2.393e-01  2.084e-02 -11.481  < 2e-16 ***
## smoke        2.908e+00  1.389e+00   2.094   0.0365 *
## trig0       -1.672e-02  1.038e-02  -1.611   0.1076
```

```
## age          9.491e-03  5.653e-02   0.168   0.8667
## gender       5.239e-01  1.268e+00   0.413   0.6797
## A:exercise  -2.108e+01  2.408e+00  -8.751  < 2e-16 ***
## A:wt         1.621e+00  2.997e-02  54.083  < 2e-16 ***
## A:smoke     -5.081e+00  1.997e+00  -2.545   0.0111 *
## A:trig0      3.512e-02  1.489e-02   2.358   0.0186 *
## A:age        2.488e-02  7.916e-02   0.314   0.7534
## A:gender     9.056e-01  1.789e+00   0.506   0.6128
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.86 on 986 degrees of freedom
## Multiple R-squared:  0.8904, Adjusted R-squared:  0.889
## F-statistic: 616.3 on 13 and 986 DF,  p-value: < 2.2e-16
```

a. regression-based estimator

```
# From slide 35 of Halloway
moMain <- buildModelObj(model = ~exercise + wt + smoke + trig0 + age + gender,
        solver.method = 'lm',
        predict.method = 'predict.lm')

moCont <- buildModelObj(model = ~exercise + wt + smoke + trig0 + age + gender,
        solver.method = 'lm',
        predict.method = 'predict.lm')

qObj <- qLearn(moMain = moMain, moCont = moCont, iter = 0L,
          data = data, response = y, txName = 'A',
          verbose = TRUE)
```

```
## First step of the Q-Learning Algorithm.
##
## Outcome regression.
## Combined outcome regression model: ~ exercise+wt+smoke+trig0+age+gender + A + A:(exercise+wt+smoke+t
## Regression analysis for Combined:
##
## Call:
## lm(formula = YinternalY ~ exercise + wt + smoke + trig0 + age +
##     gender + A + exercise:A + wt:A + smoke:A + trig0:A + age:A +
##     gender:A, data = data)
##
## Coefficients:
## (Intercept)     exercise           wt        smoke        trig0          age
##   3.246e+01    2.058e+01   -2.393e-01    2.908e+00   -1.671e-02    9.491e-03
##      gender            A   exercise:A         wt:A      smoke:A      trig0:A
##   5.239e-01   -2.615e+02   -2.108e+01    1.621e+00   -5.081e+00    3.512e-02
##       age:A     gender:A
##   2.488e-02    9.056e-01
##
##
## Recommended Treatments:
##   0   1
```

```
## 211 789
##
## Estimated value: 33.75671
```

```
coef(object = qObj)
```

```
## $outcome
## $outcome$Combined
##   (Intercept)      exercise            wt         smoke         trig0
##  3.246193e+01  2.058365e+01 -2.392622e-01  2.907925e+00 -1.671489e-02
##           age        gender             A    exercise:A          wt:A
##  9.490884e-03  5.239228e-01 -2.614569e+02 -2.107722e+01  1.620786e+00
##       smoke:A       trig0:A         age:A      gender:A
## -5.081452e+00  3.511518e-02  2.488064e-02  9.055717e-01
```

```
fitObj = fitObject(object = qObj)
fitObj
```

```
## $outcome
## $outcome$Combined
##
## Call:
## lm(formula = YinternalY ~ exercise + wt + smoke + trig0 + age +
##     gender + A + exercise:A + wt:A + smoke:A + trig0:A + age:A +
##     gender:A, data = data)
##
## Coefficients:
## (Intercept)     exercise           wt        smoke        trig0          age
##   3.246e+01    2.058e+01   -2.393e-01    2.908e+00   -1.671e-02    9.491e-03
##      gender            A   exercise:A         wt:A      smoke:A      trig0:A
##   5.239e-01   -2.615e+02   -2.108e+01    1.621e+00   -5.081e+00    3.512e-02
##       age:A     gender:A
##   2.488e-02    9.056e-01
```

```
ot <- optTx(x = qObj)
table(ot$optimalTx)
```

```
##
##   0   1
## 211 789
```

```
estimator(x = qObj)
```

```
## [1] 33.75671
```

b. restricted value search

```
# slide 54 of Halloway
regimes = function(eta1, data)
{
  d1 = {data$wt > eta1}
  return(as.integer(x = d1))
}

moPropen <- buildModelObj(model = ~ 1,
              solver.method = 'glm',
              solver.args = list(family = 'binomial'),
              predict.method = 'predict.glm',
              predict.args = list(type = 'response'))
```

c

d

```
require(rpart)
```

```
## Loading required package: rpart
```

```
moClass <- buildModelObj(model = ~exercise + wt + smoke + trig0 + age + gender,
  solver.method = 'rpart',
  predict.method = 'predict',
  predict.args = list(type = "class"))


clObj <- optimalClass(moPropen = moPropen,
            moMain = moMain, moCont = moCont, iter = 0L,
            moClass = moClass,
            data = data, response = y, txName = 'A',
            verbose = TRUE)
```

```
## AIPW value estimator

## First step of the Classification Algorithm.

## Classification Perspective.

##
## Propensity for treatment regression.
## Regression analysis for moPropen:
##
## Call:  glm(formula = YinternalY ~ 1, family = "binomial", data = data)
##
## Coefficients:
## (Intercept)
##      -0.012
##
```

```
## Degrees of Freedom: 999 Total (i.e. Null);  999 Residual
## Null Deviance:       1386
## Residual Deviance: 1386  AIC: 1388
##
## Outcome regression.
## Combined outcome regression model: ~ exercise+wt+smoke+trig0+age+gender + A + A:(exercise+wt+smoke+t
## Regression analysis for Combined:
##
## Call:
## lm(formula = YinternalY ~ exercise + wt + smoke + trig0 + age +
##     gender + A + exercise:A + wt:A + smoke:A + trig0:A + age:A +
##     gender:A, data = data)
##
## Coefficients:
## (Intercept)     exercise           wt        smoke        trig0          age
##   3.246e+01    2.058e+01   -2.393e-01    2.908e+00   -1.671e-02    9.491e-03
##      gender            A   exercise:A         wt:A      smoke:A      trig0:A
##   5.239e-01   -2.615e+02   -2.108e+01    1.621e+00   -5.081e+00    3.512e-02
##       age:A     gender:A
##   2.488e-02    9.056e-01
##
##
## Classification Analysis
## Regression analysis for moClass:
## n= 1000
##
## node), split, n, loss, yval, (yprob)
##       * denotes terminal node
##
##  1) root 1000 0.1366669000 1 (0.0244467268 0.9755532732)
##    2) wt< 158.55 227 0.0121064100 0 (0.6074944591 0.3925055409) *
##    3) wt>=158.55 773 0.0183009000 1 (0.0033918497 0.9966081503)
##      6) wt< 167.25 99 0.0139051800 1 (0.0837320373 0.9162679627)
##       12) exercise>=0.5 14 0.0006027082 0 (0.5236820030 0.4763179970) *
##       13) exercise< 0.5 85 0.0097192400 1 (0.0614852558 0.9385147442)
##         26) smoke>=0.5 24 0.0021517130 0 (0.2942207821 0.7057792179) *
##         27) smoke< 0.5 61 0.0040528840 1 (0.0291961991 0.9708038009) *
##      7) wt>=167.25 674 0.0043957250 1 (0.0008405656 0.9991594344) *
## Recommended Treatments:
##   0   1
## 265 735
##
## Estimated value: 35.1084
```

```
coef(object = clObj)
```

```
## $propensity
## (Intercept)
## -0.01200014
##
## $outcome
## $outcome$Combined
##   (Intercept)      exercise            wt         smoke         trig0
##  3.246193e+01  2.058365e+01 -2.392622e-01  2.907925e+00 -1.671489e-02
```

```
##           age        gender             A    exercise:A          wt:A
##  9.490884e-03  5.239228e-01 -2.614569e+02 -2.107722e+01  1.620786e+00
##       smoke:A       trig0:A         age:A      gender:A
## -5.081452e+00  3.511518e-02  2.488064e-02  9.055717e-01
```

```
table(ot$optimalTx)
```

```
##
##   0   1
## 211 789
```

```
estimator(x = clObj)
```

```
## [1] 35.1084
```

2

a

```
ldl = read.table("LDL.dat.txt", header=FALSE)

# remove ID column
ldl = ldl[,-1]

### Setting up variables in equations

# number of datapoints
n = dim(ldl)[1]

# number of decision points
K = 4

# LDL measurements
L = ldl[,c(1,3,6,9,12)]

# Side effect experienced
S = ldl[,c(4,7,10,13)]

# Statin dose received
A = ldl[,c(2,5,8,11)]

# Y outcome vector
# shoutout to Samsul for helping me build Y
Y = as.numeric(t(cbind(L[,1],L[,2:5]-L[,1:4])))

# X design matrix
X = NULL

for(i in 1:n){
```

```
  X = rbind(X,
            c(1, rep(0,6)))

  for(k in 2:(K+1)){
    X = rbind(X,
              c(0,
                1 - S[i, k-1],
                A[i, k-1]*(1-S[i,k-1]),
                L[i,k-1]*(1-S[i,k-1]),
                A[i,k-1]*L[i,k-1]*(1-S[i,k-1]),
                S[i,k-1],
                S[i,k-1]*A[i,k-1]))
  }
}

# fit linear model
# -1 removes intercept
lmbeta = lm(Y ~ -1 + X)
betas = coef(lmbeta)
sigmasq = (summary(lmbeta)$sigma)^2


cat("=================\n
    betas\n
    ==================")
```

```
## =================
##
##     betas
##
##     ==================
```

```
print(betas)
```

```
##            X1            X2            X3            X4            X5
## 170.092400000  -6.112302738 -11.970236677  -0.003808885   0.013909115
##            X6            X7
##  -6.592123769  -7.052320675
```

```
cat("=================\n
    sigma^2\n
    ==================")
```

```
## =================
##
##     sigma^2
##
##     ==================
```

```
print(sigmasq)
```

```
## [1] 144.3508
```

3