Birth weight case study

Subramani.M

30 May 2018

```
library(rpart)

bw_actual = read.csv("birthwt_actual.csv",header = T)
bw_tr = read.csv("birthwt_train.csv",header = T)
bw_ts = read.csv("birthwt_test.csv",header = T)

bw_tr$low = as.factor(bw_tr$low)
model = rpart(low ~ . , data = bw_tr)
model

## n= 141
##
## node), split, n, loss, yval, (yprob)
##       * denotes terminal node
##
##  1) root 141 44 0 (0.6879433 0.3120567)
##    2) ptl< 0.5 117 30 0 (0.7435897 0.2564103)
##      4) lwt>=106 93 19 0 (0.7956989 0.2043011)
##        8) ht< 0.5 86 15 0 (0.8255814 0.1744186) *
##        9) ht>=0.5 7  3 1 (0.4285714 0.5714286) *
##      5) lwt< 106 24 11 0 (0.5416667 0.4583333)
##       10) age< 22.5 15  4 0 (0.7333333 0.2666667) *
##       11) age>=22.5 9  2 1 (0.2222222 0.7777778) *
##    3) ptl>=0.5 24 10 1 (0.4166667 0.5833333)
##      6) lwt>=131.5 7  2 0 (0.7142857 0.2857143) *
##      7) lwt< 131.5 17  5 1 (0.2941176 0.7058824) *

predicted_low = predict(model,bw_ts[,-1])
predicted_low.cat <-
colnames(predicted_low)[apply(predicted_low,1,which.max)]
output <- data.frame("id" = bw_ts$id,"low" = as.integer(predicted_low.cat))
output

##    id low
## 1   1   0
## 2   2   0
## 3   3   0
## 4   4   0
## 5   5   0
## 6   6   0
## 7   7   0
## 8   8   0
```

```
## 9   9   0
## 10 10   0
## 11 21   0
## 12 22   0
## 13 23   0
## 14 24   0
## 15 25   0
## 16 26   1
## 17 27   0
## 18 28   0
## 19 29   0
## 20 30   0
## 21 31   0
## 22 32   0
## 23 33   0
## 24 34   1
## 25 35   0
## 26 36   0
## 27 37   0
## 28 38   0
## 29 39   1
## 30 40   0
## 31 41   0
## 32 42   0
## 33 43   0
## 34 44   0
## 35 45   1
## 36 46   0
## 37 47   1
## 38 48   1
```