

Analise Cardio

José Elvano Moraes

4/27/2021

```
suppressPackageStartupMessages(expr = library(tidyverse))
suppressPackageStartupMessages(expr = library(haven))
suppressPackageStartupMessages(expr = library(bnlearn))
suppressPackageStartupMessages(expr = library(Rgraphviz))
suppressPackageStartupMessages(expr = library(gRain))
suppressPackageStartupMessages(expr = library(Rmpfr))
suppressPackageStartupMessages(expr = library(readr))
suppressPackageStartupMessages(expr = library(precrec))
suppressPackageStartupMessages(expr = library(ROCR))
suppressPackageStartupMessages(expr = library(epiR))

suppressWarnings(
  cardio <- read_csv("cardio_train.csv",
    col_types = cols(id = col_integer(),
      age = col_number(),
      gender = col_factor(levels = c("1", "2")),
      height = col_number(),
      weight = col_number(),
      ap_hi = col_number(),
      ap_lo = col_number(),
      cholesterol = col_factor(levels = c("1", "2", "3")),
      gluc = col_factor(levels = c("1", "2", "3")),
      smoke = col_factor(levels = c("0", "1")),
      alco = col_factor(levels = c("0", "1")),
      active = col_factor(levels = c("0", "1")),
      cardio = col_factor(levels = c("0", "1")))
  )
)

colunas <- c("age",
  "gender",
  "height",
  "weight",
  "ap_hi",
  "ap_lo",
  "cholesterol",
  "gluc",
  "smoke",
  "alco",
  "active",
  "cardio")
```

```
dd = filter(cardio, ap_hi>0, ap_hi>ap_lo, ap_lo > 0, ap_lo<300, ap_hi>0, ap_hi<300)
set.seed(2)
```

```
ccc <- select(dd, age, height, weight, ap_lo, ap_hi)
ccc2 <- discretize(ccc, method = 'hartemink', breaks = 4)
ddf <- cbind(select(dd, -age, -height, -weight, -ap_lo, -ap_hi), ccc2)
ddf <- na.omit(ddf)
ddf$id <- NULL
```

```
amostra = ddf[1:7000,]
amostra.test = ddf[7001:nrow(dd),]
amostra.test<-na.omit(amostra.test)
summary(ddf)
```

```
##  gender   cholesterol gluc      smoke   alco      active   cardio
##  1:4780    1:5495      1:6227    0:6700   0:6951    0:1497    0:3693
##  2:2571    2:1012      2: 578    1: 651    1: 400    1:5854    1:3658
##          3: 844      3: 546
##
##          age          height      weight      ap_lo
##  (10878,17320.5] :1544   (76,159] :1936   (30,57] : 613   (10,70] :1385
##  (17320.5,19108.9]:1617   (159,164]:1631   (57,64] :1145   (70,80] :3801
##  (19108.9,21169.4]:2131   (164,170]:2371   (64,80] :3575   (80,90] :1621
##  (21169.4,23670] :2059   (170,184]:1413   (80,117]:2018   (90,100]: 544
##          ap_hi
##  (70,110] :1334
##  (110,120]:3029
##  (120,140]:2034
##  (140,180]: 954
```

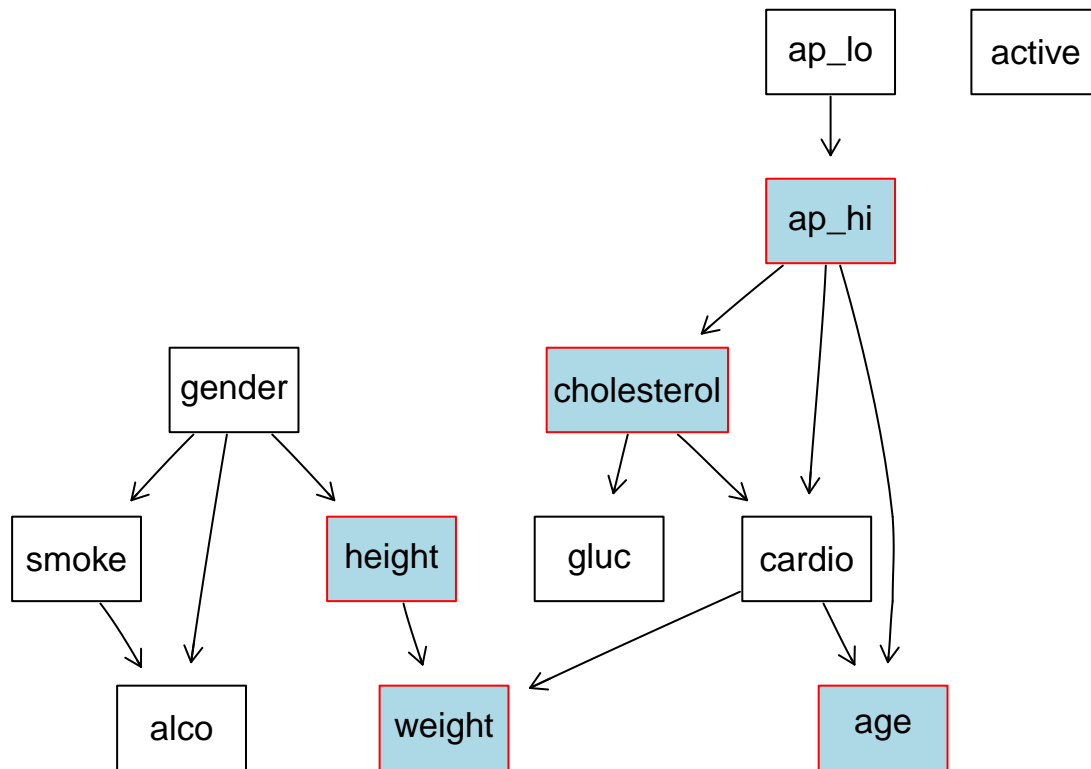
```
bn1 <- mmhc(ddf)
```

Grafo mostrando como nodos coloridos as variáveis das quais a variavel

CARDIO tem dependência estatística (*Markov blanket*)

```
# -----
graphviz.plot(bn1, shape = 'rectangle',
               main = "Markov Blanket - CARDIO",
               highlight = list(nodes=mb(bn1, 'cardio'), fill="lightblue"))
```

Markov Blanket – CARDIO



```
print("Fase de bootstrap...")
```

```
## [1] "Fase de bootstrap..."
```

```
R = 100
```

```
arc.strength = boot.strength(ddf, R = R, algorithm = "mmhc")
```

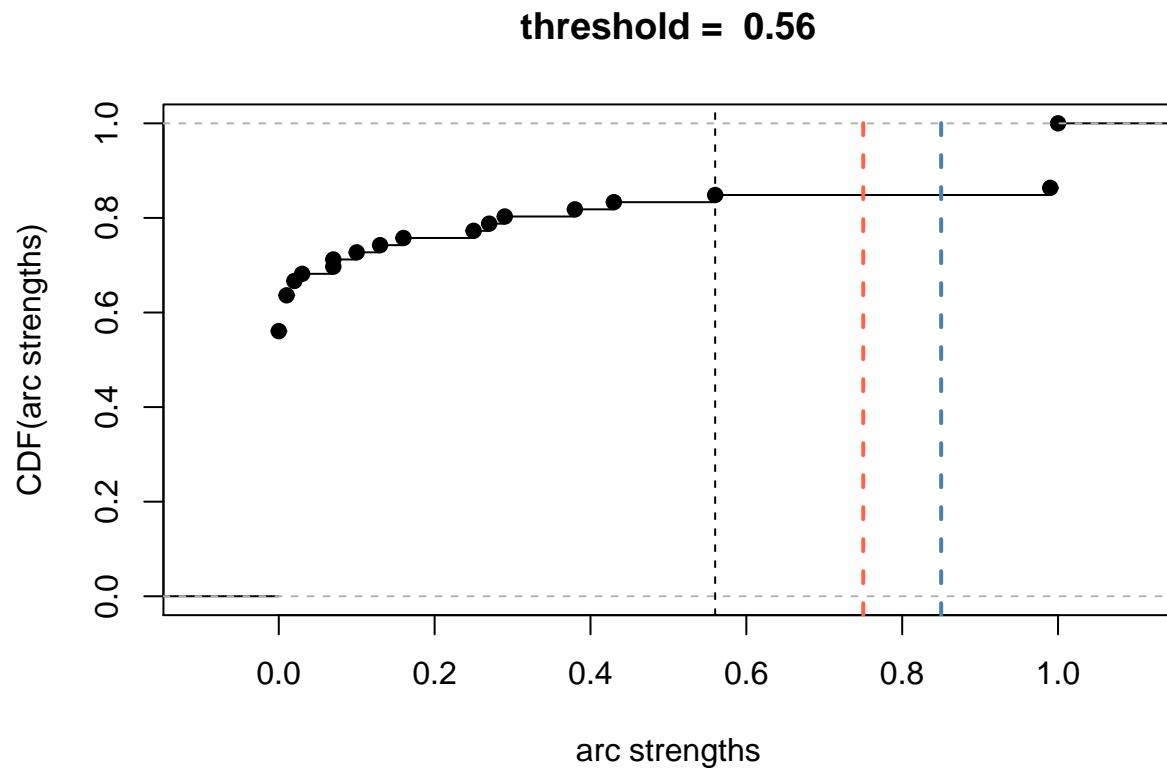
```
rede.media = averaged.network(arc.strength)
```

CDF da correlação (correlação de Pearson) entre os nodos do grafo

```
plot(arc.strength)
```

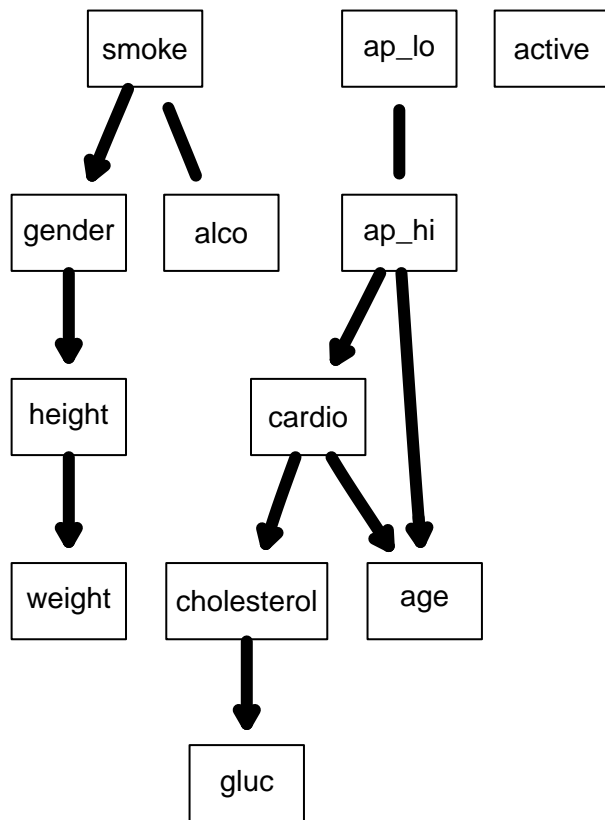
```
abline(v = 0.75, col = "tomato", lty = 2, lwd = 2)
```

```
abline(v = 0.85, col = "steelblue", lty = 2, lwd = 2)
```



Grafo médio após 100 iterações

```
par(mfrow = c(1, 2))
strength.plot(rede.media,
              arc.strength,
              shape = "rectangle")
```



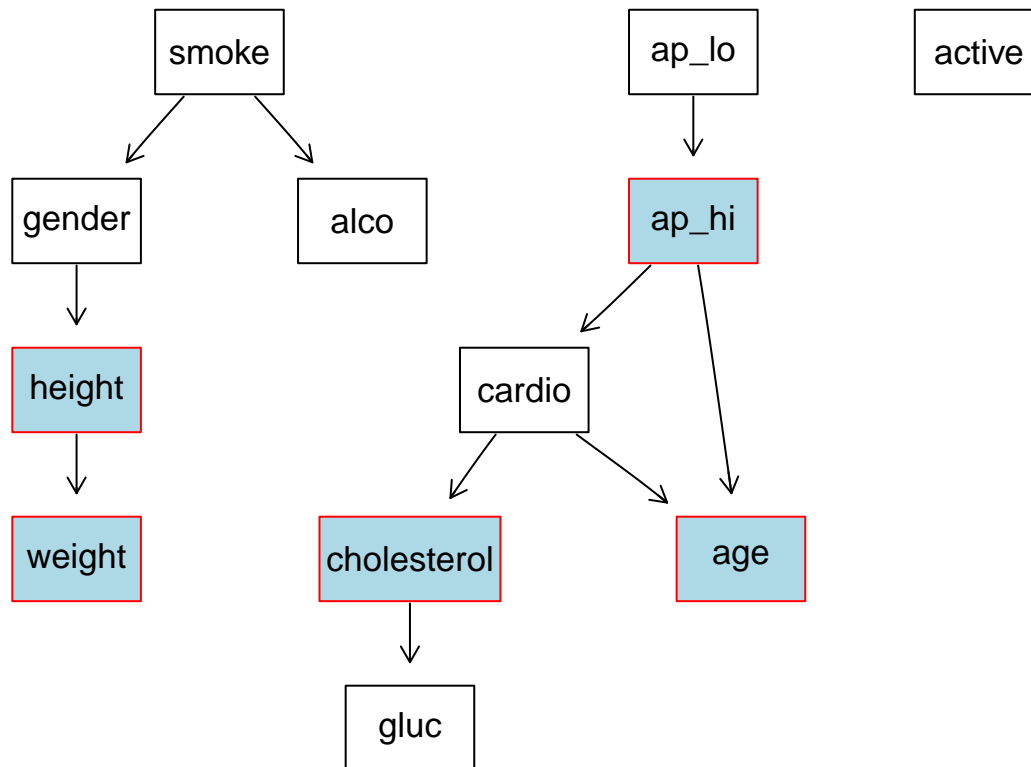
Grafo médio tornado um *DAG*

```

graphviz.plot(rede.media.dagged, shape = 'rectangle',
              main = "Markov Blanket - CARDIO",
              highlight = list(nodes=mb(bn1, 'cardio'), fill="lightblue"))

```

Markov Blanket – CARDIO

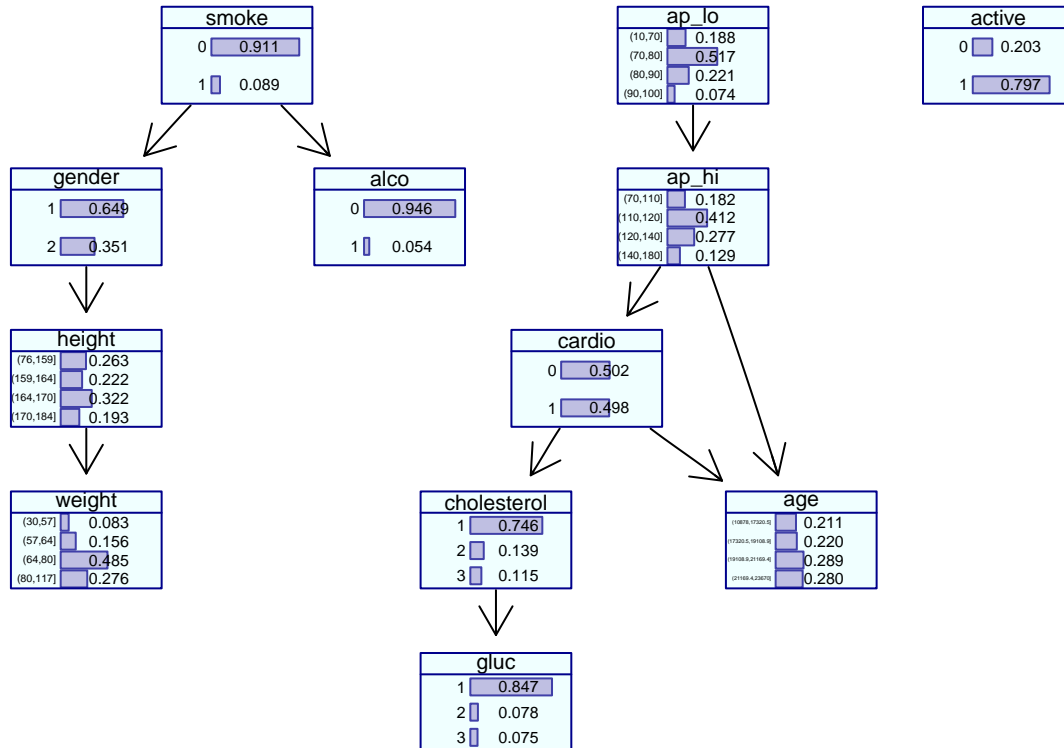


Distribuição estatística (PDF) de cada variável

```
fitted.m <- bn.fit(rede.media.dagged, amostra)

graphviz.chart(fitted.m,
  type = "barprob",
  col = "darkblue",
  bg = "azure",
  bar.col = "darkblue",
  main = "DAG sem WL")
```

DAG sem WL



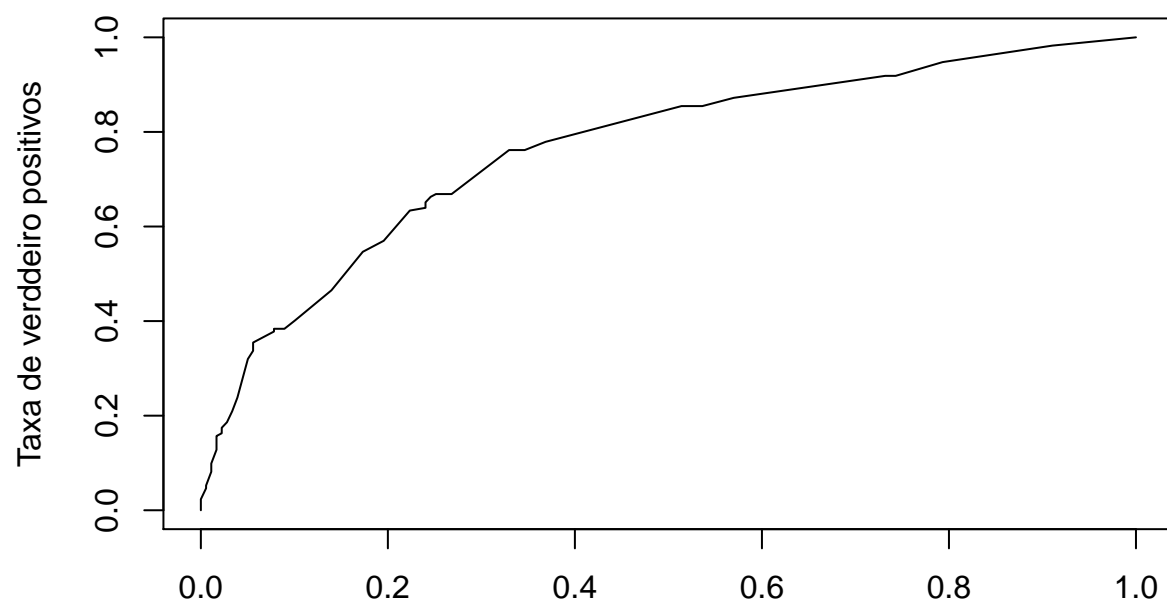
predição

Desempenho estatístico da rede bayesiana

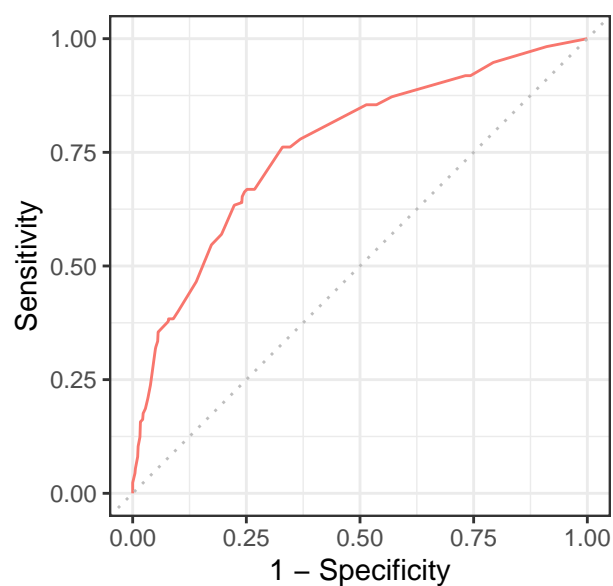
```
##          PREDICAO
## cardio    0    1
##          0 134  45
##          1  57 115

##          Outcome +   Outcome -   Total
## Test +              134          45      179
## Test -               57         115      172
## Total              191         160      351
##
## Point estimates and 95 % CIs:
## -----
## Apparent prevalence          0.51 (0.46, 0.56)
## True prevalence              0.54 (0.49, 0.60)
## Sensitivity                   0.70 (0.63, 0.77)
## Specificity                   0.72 (0.64, 0.79)
## Positive predictive value     0.75 (0.68, 0.81)
## Negative predictive value     0.67 (0.59, 0.74)
## Positive likelihood ratio     2.49 (1.91, 3.25)
## Negative likelihood ratio     0.42 (0.33, 0.53)
## -----
```

Curva ROC



ROC – P: 172, N: 179



Precision-Recall – P: 172, N: 179

