Generative Modelling to predict Heart Failure

2024-03-02

PROJECT OVERVIEW

The goal of this project is to understand heart conditions of patients, then to model the relationships between selected clinical variables, using both structural learning techniques and prior knowledge from the dataset and predicting chances of heartfailure as part of generative modelling

```r
library(bnlearn)
```

```
## Warning: package 'bnlearn' was built under R version 4.3.3
```

```r
library(gRbase)
```

```
## Warning: package 'gRbase' was built under R version 4.3.3
```

```
##
## Attaching package: 'gRbase'
```

```
## The following objects are masked from 'package:bnlearn':
##
##     ancestors, children, nodes, parents
```

```
data(cad1)

#I have decicded to use all the variables as that will help me to compare my other results

#first making the skeleton of the dag
dag2 <- empty.graph(nodes = c("Sex", "AngPec", "AMI", "QWave", "QWavecode", "STcode", "STchange", "Suff
HeartF", "Hypertrophi", "Hyperchol", "Smoker", "Inherit", "Heartfail", "CAD"))

#Now I am adding arc according to prior knowledge as per what I understood reading the help file

dag2 <- set.arc(dag2, from = "Sex", to = "AngPec") #Sex may influence AngPec as per prior knowledge
dag2 <- set.arc(dag2, from = "Hyperchol", to = "AMI")   #Hyperchol may influence AMI
dag2 <- set.arc(dag2, from = "Smoker", to = "AMI")    #Smoker may influence AMI
dag2 <- set.arc(dag2, from = "Inherit", to = "AMI")   #Inheritance may influence AMI


# Defining states for each variable
Sex_states <- c("Female", "Male")
AngPec_states <- c("Atypical", "None", "Typical")
AMI_states <- c("Definite", "NotCertain")
QWave_states <- c("No", "Yes")
QWavecode_states <- c("Nonusable", "Usable")
STcode_states <- c("Nonusable", "Usable")
STchange_states <- c("No", "Yes")
SuffHeartF_states <- c("No", "Yes")
Hypertrophi_states <- c("No", "Yes")
Hyperchol_states <- c("No", "Yes")
Smoker_states <- c("No", "Yes")
Inherit_states <- c("No", "Yes")
Heartfail_states <- c("No", "Yes")
CAD_states <- c("No", "Yes")

# Define conditional probabilities for each node

# Conditional probabilities for the Sex node
Sex_probs <- array(c(
  # P(Sex = Female)
  0.4,
  # P(Sex = Male)
  0.6
), dim = c(2), dimnames = list(Sex = Sex_states))

# Conditional probabilities for the AngPec node
AngPec_probs <- array(c(
  # P(AngPec = Atypical | Sex = Female)
  0.2,
  # P(AngPec = None | Sex = Female)
  0.6,
  # P(AngPec = Typical | Sex = Female)
  0.2,
  # P(AngPec = Atypical | Sex = Male)
  0.3,
  # P(AngPec = None | Sex = Male)
  0.5,
  # P(AngPec = Typical | Sex = Male)
  0.2
), dim = c(3, 2), dimnames = list(AngPec = AngPec_states, Sex = Sex_states))
```

```r
# Conditional probabilities for the AMI node
# Conditional probabilities for the AMI node
AMI_probs <- array(c(
  # P(AMI = Definite | Hyperchol = No, Smoker = No, Inherit = No)
  0.8,
  # P(AMI = NotCertain | Hyperchol = No, Smoker = No, Inherit = No)
  0.2,
  # P(AMI = Definite | Hyperchol = No, Smoker = Yes, Inherit = No)
  0.6,
  # P(AMI = NotCertain | Hyperchol = No, Smoker = Yes, Inherit = No)
  0.4,
  # P(AMI = Definite | Hyperchol = Yes, Smoker = No, Inherit = No)
  0.7,
  # P(AMI = NotCertain | Hyperchol = Yes, Smoker = No, Inherit = No)
  0.3,
  # P(AMI = Definite | Hyperchol = Yes, Smoker = Yes, Inherit = No)
  0.5,
  # P(AMI = NotCertain | Hyperchol = Yes, Smoker = Yes, Inherit = No)
  0.5,
  # P(AMI = Definite | Hyperchol = No, Smoker = No, Inherit = Yes)
  0.7,
  # P(AMI = NotCertain | Hyperchol = No, Smoker = No, Inherit = Yes)
  0.3,
  # P(AMI = Definite | Hyperchol = No, Smoker = Yes, Inherit = Yes)
  0.5,
  # P(AMI = NotCertain | Hyperchol = No, Smoker = Yes, Inherit = Yes)
  0.5,
  # P(AMI = Definite | Hyperchol = Yes, Smoker = No, Inherit = Yes)
  0.6,
  # P(AMI = NotCertain | Hyperchol = Yes, Smoker = No, Inherit = Yes)
  0.4,
  # P(AMI = Definite | Hyperchol = Yes, Smoker = Yes, Inherit = Yes)
  0.4,
  # P(AMI = NotCertain | Hyperchol = Yes, Smoker = Yes, Inherit = Yes)
  0.6
), dim = c(2, 2, 2, 2), dimnames = list(AMI = AMI_states, Hyperchol = Hyperchol_states, Smoker = Smoker
_states, Inherit = Inherit_states))




# Conditional probabilities for the QWave node
QWave_probs <- array(c(
  # P(QWave = No)
  0.7,
  # P(QWave = Yes)
  0.3
), dim = c(2), dimnames = list(QWave = QWave_states))

# Conditional probabilities for the QWavecode node
QWavecode_probs <- array(c(
  # P(QWavecode = Nonusable)
  0.4,
  # P(QWavecode = Usable)
  0.6
), dim = c(2), dimnames = list(QWavecode = QWavecode_states))

# Conditional probabilities for the STcode node
```

```r
STcode_probs <- array(c(
  # P(STcode = Nonusable)
  0.5,
  # P(STcode = Usable)
  0.5
), dim = c(2), dimnames = list(STcode = STcode_states))

# Conditional probabilities for the STchange node
STchange_probs <- array(c(
  # P(STchange = No)
  0.8,
  # P(STchange = Yes)
  0.2
), dim = c(2), dimnames = list(STchange = STchange_states))

# Conditional probabilities for the SuffHeartF node
SuffHeartF_probs <- array(c(
  # P(SuffHeartF = No)
  0.3,
  # P(SuffHeartF = Yes)
  0.7
), dim = c(2), dimnames = list(SuffHeartF = SuffHeartF_states))

# Conditional probabilities for the Hypertrophi node
Hypertrophi_probs <- array(c(
  # P(Hypertrophi = No)
  0.6,
  # P(Hypertrophi = Yes)
  0.4
), dim = c(2), dimnames = list(Hypertrophi = Hypertrophi_states))

# Conditional probabilities for the Hyperchol node
Hyperchol_probs <- array(c(
  # P(Hyperchol = No)
  0.7,
  # P(Hyperchol = Yes)
  0.3
), dim = c(2), dimnames = list(Hyperchol = Hyperchol_states))

# Conditional probabilities for the Smoker node
Smoker_probs <- array(c(
  # P(Smoker = No)
  0.6,
  # P(Smoker = Yes)
  0.4
), dim = c(2), dimnames = list(Smoker = Smoker_states))

# Conditional probabilities for the Inherit node
Inherit_probs <- array(c(
  # P(Inherit = No)
  0.8,
  # P(Inherit = Yes)
  0.2
), dim = c(2), dimnames = list(Inherit = Inherit_states))

# Conditional probabilities for the Heartfail node
Heartfail_probs <- array(c(
  # P(Heartfail = No)
```

```
  0.9,
  # P(Heartfail = Yes)
  0.1
), dim = c(2), dimnames = list(Heartfail = Heartfail_states))

# Conditional probabilities for the CAD node
CAD_probs <- array(c(
  # P(CAD = No)
  0.85,
  # P(CAD = Yes)
  0.15
), dim = c(2), dimnames = list(CAD = CAD_states))


#parameterizing the network
cpt <- list(Sex = Sex_probs, AngPec = AngPec_probs, AMI = AMI_probs, QWave = QWave_probs, QWavecode = Q
Wavecode_probs, STcode = STcode_probs, STchange = STchange_probs, SuffHeartF = SuffHeartF_probs, Hypert
rophi = Hypertrophi_probs, Hyperchol = Hyperchol_probs, Smoker = Smoker_probs, Inherit = Inherit_probs,
Heartfail = Heartfail_probs, CAD = CAD_probs)

bn <- custom.fit(dag2, dist = cpt)
nparams(bn)
```

```
## [1] 24
```

```
arcs(bn)
```

```
##         from          to
## [1,] "Sex"         "AngPec"
## [2,] "Hyperchol"   "AMI"
## [3,] "Smoker"      "AMI"
## [4,] "Inherit"     "AMI"
```

```
bn
```

```
##
##   Bayesian network parameters
##
##   Parameters of node Sex (multinomial distribution)
##
## Conditional probability table:
##  Sex
## Female   Male
##    0.4    0.6
##
##   Parameters of node AngPec (multinomial distribution)
##
## Conditional probability table:
##
##          Sex
## AngPec    Female Male
##   Atypical    0.2  0.3
##   None        0.6  0.5
##   Typical     0.2  0.2
##
##   Parameters of node AMI (multinomial distribution)
##
## Conditional probability table:
##
## , , Smoker = No, Inherit = No
##
##             Hyperchol
## AMI          No Yes
##   Definite   0.8 0.6
##   NotCertain 0.2 0.4
##
## , , Smoker = Yes, Inherit = No
##
##             Hyperchol
## AMI          No Yes
##   Definite   0.7 0.5
##   NotCertain 0.3 0.5
##
## , , Smoker = No, Inherit = Yes
##
##             Hyperchol
## AMI          No Yes
##   Definite   0.7 0.5
##   NotCertain 0.3 0.5
##
## , , Smoker = Yes, Inherit = Yes
##
##             Hyperchol
## AMI          No Yes
##   Definite   0.6 0.4
##   NotCertain 0.4 0.6
##
##
##   Parameters of node QWave (multinomial distribution)
##
## Conditional probability table:
##  QWave
```

```
##   No Yes
## 0.7 0.3
##
##     Parameters of node QWavecode (multinomial distribution)
##
## Conditional probability table:
##   QWavecode
## Nonusable     Usable
##        0.4        0.6
##
##     Parameters of node STcode (multinomial distribution)
##
## Conditional probability table:
##   STcode
## Nonusable     Usable
##        0.5        0.5
##
##     Parameters of node STchange (multinomial distribution)
##
## Conditional probability table:
##   STchange
##   No Yes
## 0.8 0.2
##
##     Parameters of node SuffHeartF (multinomial distribution)
##
## Conditional probability table:
##   SuffHeartF
##   No Yes
## 0.3 0.7
##
##     Parameters of node Hypertrophi (multinomial distribution)
##
## Conditional probability table:
##   Hypertrophi
##   No Yes
## 0.6 0.4
##
##     Parameters of node Hyperchol (multinomial distribution)
##
## Conditional probability table:
##   Hyperchol
##   No Yes
## 0.7 0.3
##
##     Parameters of node Smoker (multinomial distribution)
##
## Conditional probability table:
##   Smoker
##   No Yes
## 0.6 0.4
##
##     Parameters of node Inherit (multinomial distribution)
##
## Conditional probability table:
##   Inherit
##   No Yes
## 0.8 0.2
```

```
##
##   Parameters of node Heartfail (multinomial distribution)
##
## Conditional probability table:
##  Heartfail
##  No Yes
## 0.9 0.1
##
##   Parameters of node CAD (multinomial distribution)
##
## Conditional probability table:
##  CAD
##   No  Yes
## 0.85 0.15
```
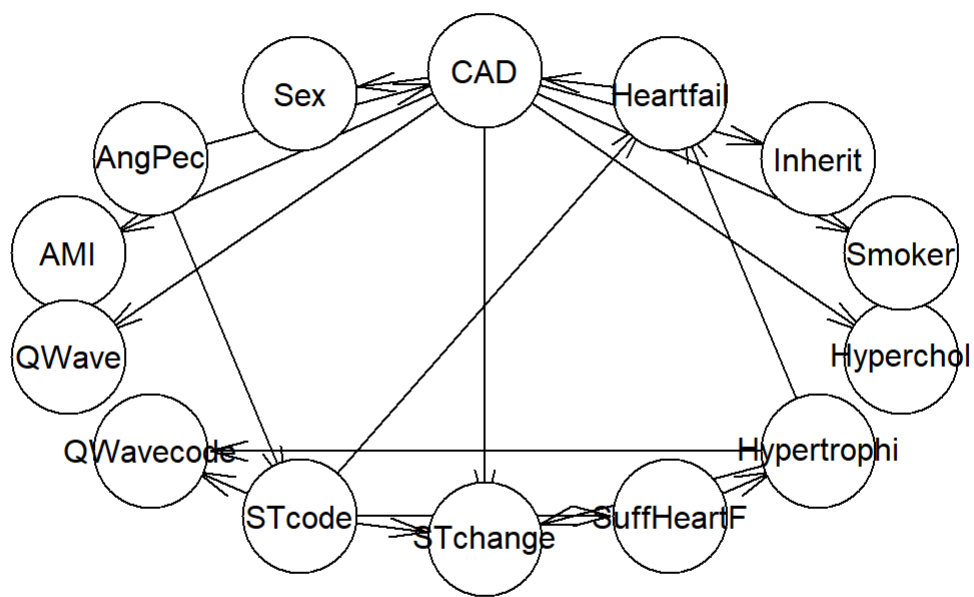
Now, infering the Conditional Probability Tables and checking for d-separations in the graph.

```
learned <- hc(cad1)
learned
```

```
##
##   Bayesian network learned via Score-based methods
##
##   model:
##    [AngPec][STcode|AngPec][SuffHeartF|STcode][Hypertrophi|SuffHeartF]
##    [QWavecode|STcode:Hypertrophi][Heartfail|STcode:Hypertrophi]
##    [CAD|AngPec:Heartfail][Sex|CAD][AMI|CAD][STchange|STcode:Hypertrophi:CAD]
##    [Hyperchol|CAD][Smoker|CAD][Inherit|CAD][QWave|AMI:CAD]
##   nodes:                                14
##   arcs:                                 19
##     undirected arcs:                    0
##     directed arcs:                      19
##   average markov blanket size:          3.29
##   average neighbourhood size:           2.71
##   average branching factor:             1.36
##
##   learning algorithm:                   Hill-Climbing
##   score:                                BIC (disc.)
##   penalization coefficient:             2.731916
##   tests used in the learning procedure: 338
##   optimized:                            TRUE
```

```
plot(learned)
```

```
#Trying several ways of finding conditional probability tables

cpt1 <- bn.fit(learned, cad1)

#Printing CPT for each node
cpt1
```

```
##
##   Bayesian network parameters
##
##   Parameters of node Sex (multinomial distribution)
##
## Conditional probability table:
##
##          CAD
## Sex             No        Yes
##   Female 0.2635659 0.1214953
##   Male   0.7364341 0.8785047
##
##   Parameters of node AngPec (multinomial distribution)
##
## Conditional probability table:
##   Atypical      None    Typical
## 0.1271186 0.3601695 0.5127119
##
##   Parameters of node AMI (multinomial distribution)
##
## Conditional probability table:
##
##            CAD
## AMI               No        Yes
##   Definite   0.09302326 0.47663551
##   NotCertain 0.90697674 0.52336449
##
##   Parameters of node QWave (multinomial distribution)
##
## Conditional probability table:
##
## , , CAD = No
##
##       AMI
## QWave  Definite NotCertain
##   No  0.4166667  0.8803419
##   Yes 0.5833333  0.1196581
##
## , , CAD = Yes
##
##       AMI
## QWave  Definite NotCertain
##   No  0.2941176  0.5357143
##   Yes 0.7058824  0.4642857
##
##
##   Parameters of node QWavecode (multinomial distribution)
##
## Conditional probability table:
##
## , , Hypertrophi = No
##
##            STcode
## QWavecode   Nonusable     Usable
##   Nonusable 0.22916667 0.01612903
##   Usable    0.77083333 0.98387097
##
```

```
## , , Hypertrophi = Yes
##
##            STcode
## QWavecode    Nonusable     Usable
##    Nonusable 0.00000000 0.00000000
##    Usable    1.00000000 1.00000000
##
##
##    Parameters of node STcode (multinomial distribution)
##
## Conditional probability table:
##
##            AngPec
## STcode        Atypical       None    Typical
##    Nonusable 0.2000000 0.5529412 0.2148760
##    Usable    0.8000000 0.4470588 0.7851240
##
##    Parameters of node STchange (multinomial distribution)
##
## Conditional probability table:
##
## , , Hypertrophi = No, CAD = No
##
##          STcode
## STchange  Nonusable     Usable
##      No   1.00000000 0.78723404
##      Yes  0.00000000 0.21276596
##
## , , Hypertrophi = Yes, CAD = No
##
##          STcode
## STchange  Nonusable     Usable
##      No   1.00000000 0.03571429
##      Yes  0.00000000 0.96428571
##
## , , Hypertrophi = No, CAD = Yes
##
##          STcode
## STchange  Nonusable     Usable
##      No   1.00000000 0.18181818
##      Yes  0.00000000 0.81818182
##
## , , Hypertrophi = Yes, CAD = Yes
##
##          STcode
## STchange  Nonusable     Usable
##      No   1.00000000 0.40000000
##      Yes  0.00000000 0.60000000
##
##
##    Parameters of node SuffHeartF (multinomial distribution)
##
## Conditional probability table:
##
##            STcode
## SuffHeartF  Nonusable     Usable
##        No   0.98734177 0.56687898
##        Yes  0.01265823 0.43312102
```

```
##
##    Parameters of node Hypertrophi (multinomial distribution)
##
## Conditional probability table:
##
##             SuffHeartF
## Hypertrophi       No        Yes
##          No   0.65269461 0.91304348
##          Yes  0.34730539 0.08695652
##
##    Parameters of node Hyperchol (multinomial distribution)
##
## Conditional probability table:
##
##            CAD
## Hyperchol       No        Yes
##       No    0.6279070 0.2523364
##       Yes   0.3720930 0.7476636
##
##    Parameters of node Smoker (multinomial distribution)
##
## Conditional probability table:
##
##        CAD
## Smoker         No        Yes
##     No   0.3100775 0.1028037
##     Yes  0.6899225 0.8971963
##
##    Parameters of node Inherit (multinomial distribution)
##
## Conditional probability table:
##
##         CAD
## Inherit        No        Yes
##     No    0.8062016 0.5420561
##     Yes   0.1937984 0.4579439
##
##    Parameters of node Heartfail (multinomial distribution)
##
## Conditional probability table:
##
## , , Hypertrophi = No
##
##           STcode
## Heartfail Nonusable    Usable
##       No   0.7500000 0.9516129
##       Yes  0.2500000 0.0483871
##
## , , Hypertrophi = Yes
##
##           STcode
## Heartfail Nonusable    Usable
##       No   0.3225806 0.3939394
##       Yes  0.6774194 0.6060606
##
##
##    Parameters of node CAD (multinomial distribution)
##
```

```
## Conditional probability table:
##
## , , Heartfail = No
##
##      AngPec
## CAD     Atypical       None    Typical
##   No  0.86363636 0.80645161 0.17204301
##   Yes 0.13636364 0.19354839 0.82795699
##
## , , Heartfail = Yes
##
##      AngPec
## CAD     Atypical       None    Typical
##   No  0.62500000 0.95652174 0.60714286
##   Yes 0.37500000 0.04347826 0.39285714
```

```
#Using Maximum Likelihood Estimation (MLE)
bn.mle <- bn.fit(dag2, data = cad1, method = "mle")

bn.mle
```

```
##
##   Bayesian network parameters
##
##   Parameters of node Sex (multinomial distribution)
##
## Conditional probability table:
##      Female      Male
## 0.1991525 0.8008475
##
##   Parameters of node AngPec (multinomial distribution)
##
## Conditional probability table:
##
##            Sex
## AngPec        Female      Male
##   Atypical 0.1276596 0.1269841
##   None     0.4893617 0.3280423
##   Typical  0.3829787 0.5449735
##
##   Parameters of node AMI (multinomial distribution)
##
## Conditional probability table:
##
## , , Smoker = No, Inherit = No
##
##            Hyperchol
## AMI               No       Yes
##   Definite   0.0800000 0.2352941
##   NotCertain 0.9200000 0.7647059
##
## , , Smoker = Yes, Inherit = No
##
##            Hyperchol
## AMI               No       Yes
##   Definite   0.1355932 0.3278689
##   NotCertain 0.8644068 0.6721311
##
## , , Smoker = No, Inherit = Yes
##
##            Hyperchol
## AMI               No       Yes
##   Definite   0.2000000 0.2500000
##   NotCertain 0.8000000 0.7500000
##
## , , Smoker = Yes, Inherit = Yes
##
##            Hyperchol
## AMI               No       Yes
##   Definite   0.3684211 0.4347826
##   NotCertain 0.6315789 0.5652174
##
##
##   Parameters of node QWave (multinomial distribution)
##
## Conditional probability table:
##        No       Yes
## 0.6483051 0.3516949
```

```
##
##   Parameters of node QWavecode (multinomial distribution)
##
## Conditional probability table:
##    Nonusable     Usable
## 0.05508475 0.94491525
##
##   Parameters of node STcode (multinomial distribution)
##
## Conditional probability table:
##  Nonusable    Usable
## 0.3347458 0.6652542
##
##   Parameters of node STchange (multinomial distribution)
##
## Conditional probability table:
##         No       Yes
## 0.5635593 0.4364407
##
##   Parameters of node SuffHeartF (multinomial distribution)
##
## Conditional probability table:
##         No       Yes
## 0.7076271 0.2923729
##
##   Parameters of node Hypertrophi (multinomial distribution)
##
## Conditional probability table:
##         No       Yes
## 0.7288136 0.2711864
##
##   Parameters of node Hyperchol (multinomial distribution)
##
## Conditional probability table:
##         No       Yes
## 0.4576271 0.5423729
##
##   Parameters of node Smoker (multinomial distribution)
##
## Conditional probability table:
##         No       Yes
## 0.2161017 0.7838983
##
##   Parameters of node Inherit (multinomial distribution)
##
## Conditional probability table:
##         No       Yes
## 0.6864407 0.3135593
##
##   Parameters of node Heartfail (multinomial distribution)
##
## Conditional probability table:
##     No  Yes
## 0.75 0.25
##
##   Parameters of node CAD (multinomial distribution)
##
## Conditional probability table:
```

```
##        No      Yes
## 0.5466102 0.4533898
```

```
bn.bayes <- bn.fit(dag2, data = cad1, method = "bayes", iss = 10)
bn.bayes
```

```
##
##   Bayesian network parameters
##
##   Parameters of node Sex (multinomial distribution)
##
## Conditional probability table:
##      Female      Male
## 0.2113821 0.7886179
##
##   Parameters of node AngPec (multinomial distribution)
##
## Conditional probability table:
##
##           Sex
## AngPec        Female      Male
##   Atypical 0.1474359 0.1323024
##   None     0.4743590 0.3281787
##   Typical  0.3782051 0.5395189
##
##   Parameters of node AMI (multinomial distribution)
##
## Conditional probability table:
##
## , , Smoker = No, Inherit = No
##
##             Hyperchol
## AMI                 No       Yes
##   Definite   0.1000000 0.2534247
##   NotCertain 0.9000000 0.7465753
##
## , , Smoker = Yes, Inherit = No
##
##             Hyperchol
## AMI                 No       Yes
##   Definite   0.1431535 0.3313253
##   NotCertain 0.8568465 0.6686747
##
## , , Smoker = No, Inherit = Yes
##
##             Hyperchol
## AMI                 No       Yes
##   Definite   0.2600000 0.3095238
##   NotCertain 0.7400000 0.6904762
##
## , , Smoker = Yes, Inherit = Yes
##
##             Hyperchol
## AMI                 No       Yes
##   Definite   0.3765432 0.4365079
##   NotCertain 0.6234568 0.5634921
##
##
##   Parameters of node QWave (multinomial distribution)
##
## Conditional probability table:
##        No       Yes
## 0.6422764 0.3577236
```

```
##
##   Parameters of node QWavecode (multinomial distribution)
##
## Conditional probability table:
##    Nonusable      Usable
## 0.07317073 0.92682927
##
##   Parameters of node STcode (multinomial distribution)
##
## Conditional probability table:
##  Nonusable     Usable
## 0.3414634 0.6585366
##
##   Parameters of node STchange (multinomial distribution)
##
## Conditional probability table:
##         No        Yes
## 0.5609756 0.4390244
##
##   Parameters of node SuffHeartF (multinomial distribution)
##
## Conditional probability table:
##        No       Yes
## 0.699187 0.300813
##
##   Parameters of node Hypertrophi (multinomial distribution)
##
## Conditional probability table:
##         No        Yes
## 0.7195122 0.2804878
##
##   Parameters of node Hyperchol (multinomial distribution)
##
## Conditional probability table:
##         No        Yes
## 0.4593496 0.5406504
##
##   Parameters of node Smoker (multinomial distribution)
##
## Conditional probability table:
##         No        Yes
## 0.2276423 0.7723577
##
##   Parameters of node Inherit (multinomial distribution)
##
## Conditional probability table:
##         No        Yes
## 0.6788618 0.3211382
##
##   Parameters of node Heartfail (multinomial distribution)
##
## Conditional probability table:
##         No        Yes
## 0.7398374 0.2601626
##
##   Parameters of node CAD (multinomial distribution)
##
## Conditional probability table:
```

```
##        No       Yes
## 0.5447154 0.4552846
```

```
# Bayesian Fit (with a prior, sample size of prior = 250)
bn.bayes2 <- bn.fit(dag2, data = cad1, method = "bayes", iss = 250)
bn.bayes2
```

```
##
##   Bayesian network parameters
##
##   Parameters of node Sex (multinomial distribution)
##
## Conditional probability table:
##     Female      Male
## 0.3539095 0.6460905
##
##   Parameters of node AngPec (multinomial distribution)
##
## Conditional probability table:
##
##           Sex
## AngPec       Female      Male
##   Atypical 0.2771318 0.2091295
##   None     0.3759690 0.3301486
##   Typical  0.3468992 0.4607219
##
##   Parameters of node AMI (multinomial distribution)
##
## Conditional probability table:
##
## , , Smoker = No, Inherit = No
##
##           Hyperchol
## AMI               No       Yes
##   Definite   0.3133333 0.4067358
##   NotCertain 0.6866667 0.5932642
##
## , , Smoker = Yes, Inherit = No
##
##           Hyperchol
## AMI               No       Yes
##   Definite   0.2617729 0.3861789
##   NotCertain 0.7382271 0.6138211
##
## , , Smoker = No, Inherit = Yes
##
##           Hyperchol
## AMI               No       Yes
##   Definite   0.4586207 0.4716312
##   NotCertain 0.5413793 0.5283688
##
## , , Smoker = Yes, Inherit = Yes
##
##           Hyperchol
## AMI               No       Yes
##   Definite   0.4502488 0.4611650
##   NotCertain 0.5497512 0.5388350
##
##
##   Parameters of node QWave (multinomial distribution)
##
## Conditional probability table:
##        No       Yes
## 0.5720165 0.4279835
```

```
##
##   Parameters of node QWavecode (multinomial distribution)
##
## Conditional probability table:
##  Nonusable    Usable
## 0.2839506 0.7160494
##
##   Parameters of node STcode (multinomial distribution)
##
## Conditional probability table:
##  Nonusable    Usable
## 0.4197531 0.5802469
##
##   Parameters of node STchange (multinomial distribution)
##
## Conditional probability table:
##         No       Yes
## 0.5308642 0.4691358
##
##   Parameters of node SuffHeartF (multinomial distribution)
##
## Conditional probability table:
##        No      Yes
## 0.600823 0.399177
##
##   Parameters of node Hypertrophi (multinomial distribution)
##
## Conditional probability table:
##         No       Yes
## 0.6111111 0.3888889
##
##   Parameters of node Hyperchol (multinomial distribution)
##
## Conditional probability table:
##         No       Yes
## 0.4794239 0.5205761
##
##   Parameters of node Smoker (multinomial distribution)
##
## Conditional probability table:
##         No       Yes
## 0.3621399 0.6378601
##
##   Parameters of node Inherit (multinomial distribution)
##
## Conditional probability table:
##        No      Yes
## 0.590535 0.409465
##
##   Parameters of node Heartfail (multinomial distribution)
##
## Conditional probability table:
##         No       Yes
## 0.6213992 0.3786008
##
##   Parameters of node CAD (multinomial distribution)
##
## Conditional probability table:
```

```
##        No       Yes
## 0.5226337 0.4773663
```

```
#Now checking for d-separations

#checking if CAD and AMI are d-separated
dsep(dag2, x ="CAD", y= "AMI")
```

```
## [1] TRUE
```

```
#Checking if CAD and AngPec are d-separated given evidence on Smoker
dsep(dag2, x = "CAD", y = "AngPec", z = "Smoker")
```

```
## [1] TRUE
```

```
#Checking if SuffHeartF and Hypertrophi are d-separated given evidence on CAD
dsep(dag2, x = "SuffHeartF", y = "Hypertrophi", z = "CAD")
```

```
## [1] TRUE
```

```
#Checking if Heartfail and Smoker are d-separated given evidence on CAD and AMI
dsep(dag2, x = "Heartfail", y = "Smoker", z = c("CAD", "AMI"))
```

```
## [1] TRUE
```

```
#Checking if Sex and Inherit are d-separated given no additional evidence
dsep(dag2, x = "Sex", y = "Inherit")
```

```
## [1] TRUE
```

#All of the above conditions return True so they are deseparated. #This means CAD and AMI are d-separated, CAD and AngPec are d-separated given evidence on Smoker, SuffHeartF and Hypertrophi are d-separated given evidence on CAD, Heartfail and Smoker are d-separated given evidence on CAD and AMI and Sex and Inherit are d-separated given no additional evidence.

Now for cases when a female has Hypercholesterolemia (high cholesterol), lets predict the probability of heart-failure and coronary artery disease (CAD).

```
library(gRain)
```

```
## Warning: package 'gRain' was built under R version 4.3.3
```

```
learned <- hc(cad1)
learned
```

```
##
##    Bayesian network learned via Score-based methods
##
##    model:
##     [AngPec][STcode|AngPec][SuffHeartF|STcode][Hypertrophi|SuffHeartF]
##     [QWavecode|STcode:Hypertrophi][Heartfail|STcode:Hypertrophi]
##     [CAD|AngPec:Heartfail][Sex|CAD][AMI|CAD][STchange|STcode:Hypertrophi:CAD]
##     [Hyperchol|CAD][Smoker|CAD][Inherit|CAD][QWave|AMI:CAD]
##    nodes:                              14
##    arcs:                               19
##      undirected arcs:                  0
##      directed arcs:                    19
##    average markov blanket size:        3.29
##    average neighbourhood size:         2.71
##    average branching factor:           1.36
##
##    learning algorithm:                 Hill-Climbing
##    score:                              BIC (disc.)
##    penalization coefficient:           2.731916
##    tests used in the learning procedure: 338
##    optimized:                          TRUE
```

```
junction <- compile(as.grain(cpt1))

# Plot the Bayesian network
plot(junction)
```
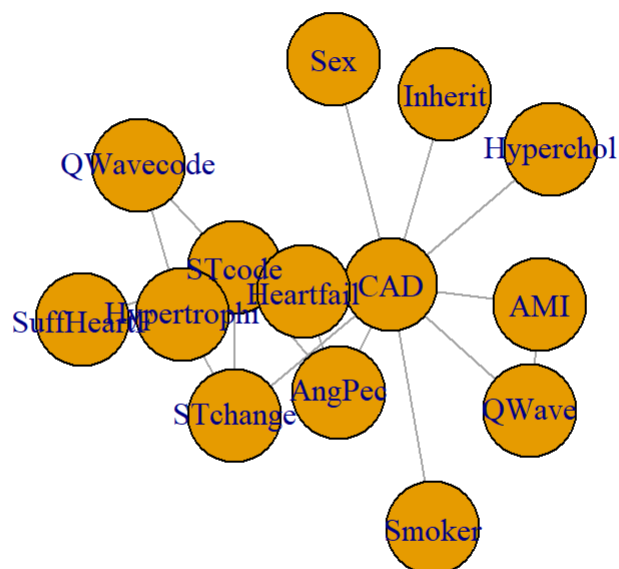
```
# Exact inference with evidence
jedu <- setEvidence(junction, nodes = c("Sex", "Hyperchol"), states = c("Female", "High"))
SxT <- querygrain(jedu, nodes = c("HeartFailure", "CAD"), type = "joint")
SxT
```

```
## CAD
##         No        Yes
## 0.7247719 0.2752281
```

cpt1$Heartfail

```
##
##    Parameters of node Heartfail (multinomial distribution)
##
## Conditional probability table:
##
## , , Hypertrophi = No
##
##          STcode
## Heartfail Nonusable    Usable
##       No  0.7500000 0.9516129
##      Yes  0.2500000 0.0483871
##
## , , Hypertrophi = Yes
##
##          STcode
## Heartfail Nonusable    Usable
##       No  0.3225806 0.3939394
##      Yes  0.6774194 0.6060606
```

cpt1$CAD

```
##
##    Parameters of node CAD (multinomial distribution)
##
## Conditional probability table:
##
## , , Heartfail = No
##
##      AngPec
## CAD      Atypical       None    Typical
##    No  0.86363636 0.80645161 0.17204301
##   Yes  0.13636364 0.19354839 0.82795699
##
## , , Heartfail = Yes
##
##      AngPec
## CAD      Atypical       None    Typical
##    No  0.62500000 0.95652174 0.60714286
##   Yes  0.37500000 0.04347826 0.39285714
```

#changes in the probability of heart-failure and coronary artery disease (CAD) after this information is considered

Given the Bayesian network structure learned via Hill-Climbing and the evidence provided, we can perform inference to obtain the updated probabilities of Heartfail and CAD.

Before Evidence We had the probabilities of Heartfail and CAD based on the Bayesian network structure and possibly some prior data.

After Evidence By incorporating the evidence that the new observation is female with hypercholesterolemia, the probabilities of Heartfail and CAD are expected to change. The probability of Heartfail and CAD will be updated based on the new evidence and the dependencies specified in the Bayesian network. Therefore, the probability of CAD may change due to the new evidence about hypercholesterolemia. Given the new evidence of being female with hypercholesterolemia, the probabilities of Heartfail and CAD will likely be adjusted based on how these variables interact with the observed evidence and other factors in the network. For example:

The probability of Heartfail may increase if hypercholesterolemia (Hyperchol) is associated with a higher likelihood of heart failure. Similarly, the probability of CAD may also increase if hypercholesterolemia is linked to a higher risk of coronary artery disease in the network. However, the exact changes in the probabilities would depend on the strength of the evidence provided by the new observation and the conditional dependencies specified in the Bayesian network. To determine the precise changes, we would need to perform inference with evidence in the Bayesian network, incorporating the new evidence of being female with hypercholesterolemia and considering its interactions with other variables in the network.

Therefore, the exact change in the probabilities of Heartfail and CAD can be obtained by performing inference with the evidence provided. We can query the joint distribution of Heartfail and CAD given the evidence to see how their probabilities change after considering the new information about the new observation being female with hypercholesterolemia. By performing inference with evidence in the Bayesian network, we can obtain the updated probabilities of Heartfail and CAD, providing insights into how the new information affects these variables in the context of the learned network structure.

Data and Variables Selection

I began by utilizing the cad1 dataset from the bnlearn package, which contains various clinical variables pertinent to heart disease and related conditions. To create a comprehensive model of heart failure and CAD, I included all available variables, such as sex, smoking habits, cholesterol levels, heart failure, and several other clinical indicators. By using a broad set of variables, I aimed to capture a wide range of relationships, allowing for more thorough comparisons between different models and outcomes.

Structural Learning: Building the DAG

Step 1: Initializing the Graph The first step in constructing the Bayesian network was to create the skeleton of a directed acyclic graph (DAG). In this graph, each node represented a clinical variable, and the edges (arcs) between nodes indicated potential causal relationships.

Step 2: Incorporating Prior Knowledge To inform the structure of the DAG, I integrated prior knowledge about the relationships between variables. For example, I assumed that factors like smoking, hypercholesterolemia, and genetic inheritance could influence the likelihood of acute myocardial infarction (AMI). Similarly, I posited that sex could affect the likelihood of developing angina pectoris (AngPec). These assumptions were used to add directed arcs between the relevant nodes in the DAG.

This process of incorporating prior knowledge was crucial for capturing realistic dependencies between clinical factors, reflecting how lifestyle choices and genetics contribute to heart disease outcomes.

Defining Conditional Probabilities

With the DAG structure in place, the next step was to define the conditional probability tables (CPTs) for each node in the network. These CPTs specify the probability distribution of each variable, given the states of its parent variables.

Step 1: Setting States For each variable, I defined a set of discrete states. For example, "Sex" had states "Male" and "Female," while "Smoker" had states "Yes" and "No." These states allowed me to model binary or categorical distinctions between different clinical factors.

Step 2: Defining Conditional Probability Distributions I then specified the actual probability distributions for each variable. For instance, I defined the probabilities of different types of angina (AngPec) based on the patient's sex, and I determined the probability of AMI based on the combination of hypercholesterolemia, smoking, and genetic inheritance.

These conditional probabilities allowed me to parameterize the Bayesian network, enabling it to model how changes in one variable influence others within the system.

Learning the Network Structure

In addition to incorporating prior knowledge, I employed a data-driven approach to learn the network structure using the Hill-Climbing (HC) algorithm. This algorithm iteratively searches for the best structure by optimizing the fit between the model and the data.

Comparison of Fitting Methods

To enhance the model, I applied different fitting methods:

Maximum Likelihood Estimation (MLE): This method estimates the most likely parameter values based on the observed data. Bayesian Estimation: I used Bayesian estimation with different priors to account for uncertainty and incorporate prior beliefs about the relationships between variables, adjusting these beliefs based on the data. D-Separation and Conditional Independence I conducted d-separation tests to determine if certain variables in the network were conditionally independent given other variables. D-separation is a key concept in Bayesian networks that helps identify which variables are independent of each other under different conditions.

For instance:

CAD and AMI: I checked if these two variables were d-separated, indicating conditional independence. CAD and Angina (AngPec): I tested whether these variables were d-separated given evidence of smoking. The tests showed that these variables were d-separated in all cases, confirming the conditional independence of these variables in the network structure.

Inference with Evidence

With the Bayesian network constructed and fitted to the data, I performed inference to predict heart failure and CAD probabilities based on new evidence, specifically considering a female patient with hypercholesterolemia.

Step 1: Initial Probabilities Initially, I examined the baseline probabilities of heart failure and CAD without incorporating any new evidence.

Step 2: Updating Probabilities with Evidence Next, I introduced evidence that the patient was female and had high cholesterol. Using this information, I updated the joint probabilities of heart failure and CAD through exact inference in the Bayesian network.

Step 3: Change in Probabilities The inference revealed that the probabilities of heart failure and CAD increased when the new evidence was considered. This result demonstrates the utility of Bayesian networks in dynamically updating predictions based on new information, making them highly valuable for real-time decision-making in clinical settings.

Conclusion

In this project, I successfully built a Bayesian network to model the relationships between various clinical variables associated with heart failure and CAD. By combining prior knowledge with data-driven techniques, I was able to create a well-parameterized network, validate its structure through d-separation tests, and perform inference to predict outcomes based on new evidence.

This work highlights the effectiveness of generative models like Bayesian networks in understanding complex medical conditions and provides a robust framework for predicting heart disease outcomes in clinical practice. The ability to update predictions based on new evidence makes these models particularly powerful for dynamic and personalized medical decision-making.