



Bayesian reasoning

Laboratory activity

Project title: Heart failure prediction

Name: Vasile-Laurentiu Campean
Group: 30231
Email: laurentiucampean@gmail.com



Contents

1	Crearea rețelei Bayesiane	3
1.1	Problema propusa si modelarea rețelei	3
1.2	Lista variabilelor si continutul lor	3
2	Rezultate	6
3	Own code	8

Chapter 1

Crearea rețelei Bayesiane

1.1 Problema propusa si modelarea rețelei

Dorim sa cream o retea Bayesiană, care sa calculeze riscul mortii unei persoane in cazul unor boli asociate cu afectiunile pe inima. De asemenea, vom putea vedea si probabilitatea unei persoane de a deceda, avand in vedere mai multi factori, care cuprind atat attribute despre aceasta (varsta, sex), cat si faptul daca aceste persoana sufera sau nu de anumite boli.

1.2 Lista variabilelor si continutul lor

- Variabila Smoker are domeniul "True" sau "False" si reprezinta probabilitatea ca o persoana sa fie fumatoare.
- Variabila Age are domeniul "<35" respectiv ">35" si reprezinta probabilitatea ca o persoana sa aiba varsta mai mica de 35 de ani, sau mai mare.
- Variabila Gender are domeniul "Male" sau "Female" si reprezinta probabilitatea ca o persoana sa fie barbat sau femeie.
- Variabila Hypertension are domeniul "True" sau "False" si reprezinta probabilitatea ca o persoana sa fie sau nu hipertensiva. Aceasta este influentata de variabilele: Smoker, Age si Gender.
- Variabila Diabetes are domeniul "True" sau "False" si reprezinta probabilitatea ca o persoana sa fie sau nu diabetica. Aceasta este influentata de variabilele Age si Gender.
- Variabila Anaemia are domeniul "True" sau "False" si reprezinta probabilitatea ca o persoana sa fie sau nu anemica. Aceasta este influentata de variabilele Diabetes si Gender.
- Variabila Death are domeniul "True" sau "False" si reprezinta probabilitatea ca o persoana sa decedeze. Aceasta este influentata de variabilele: Hypertension, Diabetes si Anaemia.

Probabilitatea de a fi fumator:

Smoker	
True	0.75
False	0.25

Probabilitatea de a fi barbat sau femeie:

Gender	
Male	0.51
Female	0.49

Probabilitatea de a avea varsta mai mica sau mai mare de 35 de ani:

Age	
<35	0.47
>35	0.53

Probabilitatea de a avea hipertensiune:

Age	Gender	Smoker	$P(\text{Hypertension} = T)$	$P(\text{Hypertension} = F)$
<35	Female	True	0.31	0.69
<35	Female	False	0.24	0.76
<35	Male	True	0.32	0.68
<35	Male	False	0.27	0.73
>35	Female	True	0.39	0.61
>35	Female	False	0.34	0.66
>35	Male	True	0.33	0.67
>35	Male	False	0.3	0.7

Probabilitatea de a avea diabet:

Age	Gender	$P(\text{Diabetes} = T)$	$P(\text{Diabetes} = F)$
<35	Female	0.21	0.79
<35	Male	0.27	0.73
>35	Female	0.39	0.61
>35	Male	0.45	0.55

Probabilitatea de a fi anemic:

Gender	Diabetes	$P(\text{Anaemia} = T)$	$P(\text{Anaemia} = F)$
Female	True	0.38	0.62
Female	False	0.16	0.84
Male	True	0.21	0.79
Male	False	0.11	0.89

Probabilitatea de a deceda:

Anaemia	Diabetes	Hypertension	$P(\text{Death} = T)$	$P(\text{Death} = F)$
True	True	True	0.97	0.3
True	True	False	0.65	0.35
True	False	True	0.6	0.4
True	False	False	0.3	0.7
False	True	True	0.55	0.45
False	True	False	0.3	0.7
False	False	True	0.25	0.75
False	False	False	0.03	0.97

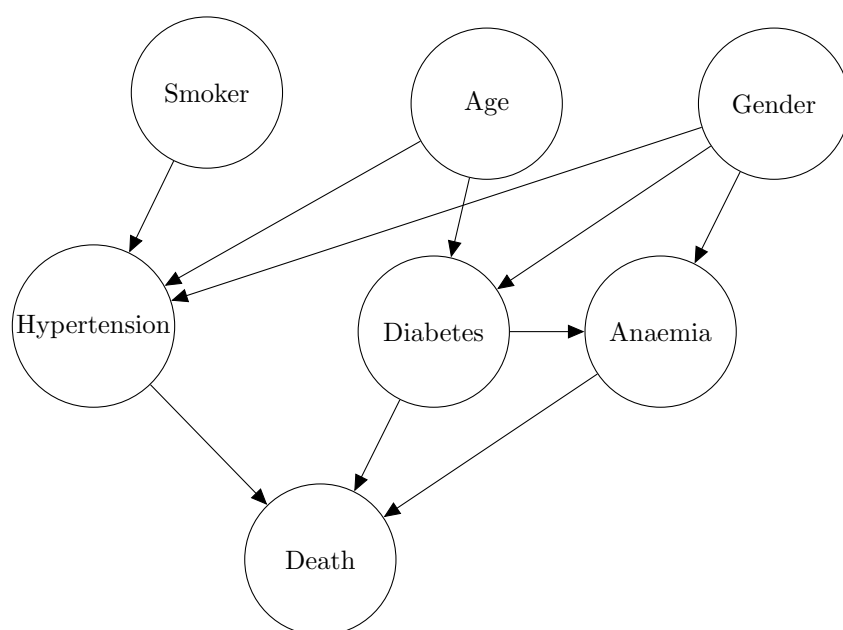


Figure 1.1: Reteaua Bayesiană și legăturile dintre noduri

Chapter 2

Rezultate

Rezultate interogari retea Bayesiană

Am ales sa prezint cateva interogari care sa calculeze probabilitatea unei persoane de a deceda sau nu, in functie de caracteristicile ei:

Probabilitatea de a deceda, stiind ca: esti fumator, >35 ani, esti barbat, ai toate cele 3 boli:

$$P(Smoker = True \wedge Age > 35 \wedge Gender = Male \wedge Diabetes = True \wedge Hypertension = True \wedge Anaemia = True \wedge Death = True) = 0.25 * 0.53 * 0.51 * 0.33 * 0.45 * 0.21 * 0.97 = 0.00204410658$$

Probabilitatea de a nu deceda, stiind ca: nu esti fumator, >35 ani, esti barbat, nu ai nicio boala:

$$P(Smoker = False \wedge Age > 35 \wedge Gender = Male \wedge Diabetes = False \wedge Hypertension = False \wedge Anaemia = False \wedge Death = False) = 0.75 * 0.53 * 0.51 * 0.7 * 0.55 * 0.89 * 0.97 = 0.0673798096$$

Probabilitatea de a nu avea diabet, stiind ca esti femeie:

$$P(Diabetes = T | Female = T) = (P((Female = True | Age < 35) | Diabetes) + P((Female = True | Age > 35) | Diabetes)) * P(Gender = Female) = (0.21 + 0.39) * 0.49 \approx 0.294$$

```
        "frozen" : false
    }
    Diabetes
    {
        "class" : "Distribution",
        "dtype" : "str",
        "name" : "DiscreteDistribution",
        "parameters" : [
            {
                "False" : 0.7284499360304734,
                "True" : 0.2715500639695265
            }
        ],
        "frozen" : false
    }
    Anaemia
    {
```

Probabilitatea de a fi anemic, știind ca nu esti diabetic:

$$P(\text{Anaemia} = T | \text{Diabetes} = F) = (P(\text{Gender} = \text{Male} | \text{Diabetes} = \text{False}) | \text{Anaemia}) * P(\text{Gender} = \text{Male}) + (P(\text{Gender} = \text{Female} | \text{Diabetes} = \text{False}) | \text{Anaemia}) * P(\text{Gender} = \text{Female}) = 0.11 * 0.51 + 0.16 * 0.49 \approx 0.1345$$

```
Diabetes          False
Anaemia          {
  "class" : "Distribution",
  "dtype" : "str",
  "name" : "DiscreteDistribution",
  "parameters" : [
    {
      "False" : 0.883484570297497,
      "True" : 0.11651542970250302
    }
  ],
  "frozen" : false
}
```

Chapter 3

Own code

```
1  #!/usr/bin/env python
2  # coding: utf-8
3
4  # In[2]:
5
6
7  get_ipython().run_line_magic('matplotlib', 'inline')
8  import matplotlib.pyplot as plt
9  import seaborn; seaborn.set_style('whitegrid')
10 import numpy as np
11
12 from pomegranate import *
13
14 numpy.random.seed(0)
15 numpy.set_printoptions(suppress=True)
16
17 get_ipython().run_line_magic('load_ext', 'watermark')
18 get_ipython().run_line_magic('watermark', '-m -n -p numpy,scipy,pomegranate'
19 )
20
21 # In[19]:
22
23
24 Smoker = DiscreteDistribution({ 'True': 0.25 , 'False': 0.75})
25 Age = DiscreteDistribution({ 'Less35': 0.47 , 'More35': 0.53})
26 Gender = DiscreteDistribution({ 'F': 0.49 , 'M': 0.51})
27
28
29 # In[34]:
30
31
32 Hypertension = ConditionalProbabilityTable(
33     [[ 'Less35', 'F', 'True', 'True', 0.31 ],
34      [ 'Less35', 'F', 'False', 'True', 0.24 ],
35      [ 'Less35', 'M', 'True', 'True', 0.32 ],
36      [ 'Less35', 'M', 'False', 'True', 0.27 ],
37      [ 'More35', 'F', 'True', 'True', 0.39 ],
38      [ 'More35', 'F', 'False', 'True', 0.34 ],
39      [ 'More35', 'M', 'True', 'True', 0.33 ],
40      [ 'More35', 'M', 'False', 'True', 0.3 ],
41      [ 'Less35', 'F', 'True', 'False', 0.69 ],
42      [ 'Less35', 'F', 'False', 'False', 0.76 ],
43      [ 'Less35', 'M', 'True', 'False', 0.68 ],
44      [ 'Less35', 'M', 'False', 'False', 0.73 ],
```



```

45     [ 'More35', 'F', 'True', 'False', 0.61 ],
46     [ 'More35', 'F', 'False', 'False', 0.66 ],
47     [ 'More35', 'M', 'True', 'False', 0.67 ],
48     [ 'More35', 'M', 'False', 'False', 0.7 ]], [Age, Gender, Smoker] )
49
50 Diabetes = ConditionalProbabilityTable([[ 'Less35', 'F', 'True', 0.21 ],
51                                         [ 'Less35', 'M', 'True', 0.27 ],
52                                         [ 'More35', 'F', 'True', 0.39 ],
53                                         [ 'More35', 'M', 'True', 0.45 ],
54                                         [ 'Less35', 'F', 'False', 0.79],
55                                         [ 'Less35', 'M', 'False', 0.73 ],
56                                         [ 'More35', 'F', 'False', 0.61 ],
57                                         [ 'More35', 'M', 'False', 0.55 ]],
58                                         [Age, Gender])
59 Anaemia = ConditionalProbabilityTable([[ 'F', 'True', 'True', 0.38 ],
60                                         [ 'F', 'False', 'True', 0.16 ],
61                                         [ 'M', 'True', 'True', 0.21 ],
62                                         [ 'M', 'False', 'True', 0.11 ],
63                                         [ 'F', 'True', 'False', 0.62],
64                                         [ 'F', 'False', 'False', 0.84 ],
65                                         [ 'M', 'True', 'False', 0.79 ],
66                                         [ 'M', 'False', 'False', 0.89 ]],
67                                         [Gender, Diabetes])
68
69 Death = ConditionalProbabilityTable(
70     [[ 'True', 'True', 'True', 'True', 0.97 ],
71       [ 'True', 'True', 'False', 'True', 0.65 ],
72       [ 'True', 'False', 'True', 'True', 0.6 ],
73       [ 'True', 'False', 'False', 'True', 0.3],
74       [ 'False', 'True', 'True', 'True', 0.55 ],
75       [ 'False', 'True', 'False', 'True', 0.3 ],
76       [ 'False', 'False', 'True', 'True', 0.25 ],
77       [ 'False', 'False', 'False', 'True', 0.3 ],
78       [ 'True', 'True', 'True', 'False', 0.03 ],
79       [ 'True', 'True', 'False', 'False', 0.35 ],
80       [ 'True', 'False', 'True', 'False', 0.4 ],
81       [ 'True', 'False', 'False', 'False', 0.7 ],
82       [ 'False', 'True', 'True', 'False', 0.45],
83       [ 'False', 'True', 'False', 'False', 0.7],
84       [ 'False', 'False', 'True', 'False', 0.75 ],
85       [ 'False', 'False', 'False', 'False', 0.97 ]], [Anaemia, Diabetes,
86       Hypertension] )
87
88 # In[35]:
89
90
91 s0 = State( Smoker, name="Smoker" )
92 s1 = State( Age, name="Age" )
93 s2 = State( Gender, name="Gender" )
94 s3 = State( Hypertension, name="Hypertension" )
95 s4 = State( Diabetes, name="Diabetes")
96 s5 = State(Anaemia, name="Anaemia")
97 s6 = State(Death, name="Death")
98
99
100 # In[36]:
101
102
103 network = BayesianNetwork( "Health" )

```

```

104 network.add_nodes(s0, s1, s2, s3, s4, s5, s6)
105
106 network.add_edge(s0, s3)
107 network.add_edge(s1, s3)
108 network.add_edge(s2, s3)
109
110 network.add_edge(s1, s4)
111 network.add_edge(s2, s4)
112
113 network.add_edge(s2, s5)
114 network.add_edge(s4, s5)
115
116 network.add_edge(s5, s6)
117 network.add_edge(s4, s6)
118 network.add_edge(s3, s6)
119
120
121 # In[37]:
122
123
124 network.bake()
125
126
127 # In[38]:
128
129
130 print (network.probability(np.array(['True', 'More35', 'M', 'True', 'True',
131                                     'True', 'True' ] , ndmin =2)))
132 #Probabilitatea de a deceda, stiind ca: esti fumator, > 35 ani, esti barbat,
133     ai toate cele 3 boli
134 #0.0020441065837499997
135
136
137 # In[39]:
138
139
140 print (network.probability(np.array(['False', 'More35', 'M', 'False', 'False
141                                     ', 'False', 'False' ] , ndmin =2)))
142 #Probabilitatea de a nu deceda, stiind ca: nu esti fumator, > 35 ani, esti
143     barbat, nu ai nicio boala.
144 #0.0673798096125
145
146
147 # In[51]:
148
149
150 observations = { 'Gender' : 'F'}
151 beliefs = map( str, network.predict_proba( observations ) )
152 print("\n".join( "{}\t\t{}".format( state.name, belief ) for state, belief
153                                     in zip( network.states, beliefs ) ))
154 #folositi pentru a vedea probabilitatea de a nu avea diabet, fiind femeie
155
156
157 # In[52]:
158
159
160 observations = { 'Diabetes' : 'False'}
161 beliefs = map( str, network.predict_proba( observations ) )
162 print("\n".join( "{}\t\t{}".format( state.name, belief ) for state, belief
163                                     in zip( network.states, beliefs ) ))

```

```
158 #Folosit pentru a vedea probabilitatea de a fi anemic, stiind ca nu esti  
    diabetic
```

Listing 3.1: Code for the Heart Failure Network

Bibliography

https://www.ined.fr/en/everything_about_population/demographic-facts-sheets/faq/more-men-or-women-in-the-world/
<https://www.kaggle.com/andrewmvd/heart-failure-clinical-data>

Intelligent Systems Group

