UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE TECNOLOGIA – CT
DEPARTAMENTO DE COMPUTAÇÃO E AUTOMAÇÃO
DCA0200 – INTELIGÊNCIA ARTIFICIAL (2023.1 - T01)

Alunos: José Augusto e Luiz Henrique

Lista de Exercícios – 1ª Unidade

1) A tabela de dados abaixo ilustra a aplicação do método Naïve-Bayes. Um determinado banco deve decidir se um cliente deve ou não receber um empréstimo bancário em função da sua condição de bom ou mau pagador. Considerando os dados de treinamento abaixo, aplique o classificador Naive-Bayes, para atribuir a classe (rótulo) para os registros 12 e 13:

Registro	Tem casa própria	Estado Civil	Possui Carro	Rendimentos	Bom Pagador
1	Sim	Solteiro	Sim	Alto	Sim
2	Não	Casado	Sim	Médio	Não
3	Não	Solteiro	Não	Baixo	Não
4	Sim	Casado	Sim	Alto	Não
5	Não	Divorciado	Não	Médio	Sim
6	Não	Casado	Não	Baixo	Não
7	Sim	Divorciado	Sim	Alto	Sim
8	Não	Solteiro	Sim	Médio	Sim
9	Não	Casado	Sim	Baixo	Não
10	Não	Solteiro	Não	Médio	Sim
11	Sim	Divorciado	Não	Médio	Não
12	Não	Divorciado	Sim	Alto	?
13	Sim	Solteiro	Não	Médio	?

Para atribuir se é ou não um bom pagador ao registro 12, é preciso fazer dois cálculos de probabilidade e optar pelo que gerar um maior resultado. O que irá determinar essas probabilidades é o Teorema de Naive Bayes, que é descrito pela equação a seguir:

$$P(H \mid E) = \frac{P(H) * \prod_{i=1}^{N} P(e_i \mid H)}{\sum_{j=1}^{K} P(E \mid h = h_j) * P(h = h_j)}$$

- P(Bom_Pagador = Sim | Tem_Casa_Própria = Não, Estado_Civil = Divorciado, Possui_Carro = Sim, Rendimentos = Alto) =

$$\frac{P(TCP=N \mid BP=S)*P(EC=D \mid BP=S)*P(PC=S \mid BP=S)*P(R=A \mid BP=S)*P(BP=S)}{P(TCP=N \mid BP=S)*P(EC=D \mid BP=S)*P(PC=S \mid BP=S)*P(R=A \mid BP=S)*P(BP=S)+P(TCP=N \mid BP=N)*P(EC=D \mid BP=N)*P(PC=S \mid BP=N)*P(R=A \mid BP=N)*P(BP=N)}$$

- P (Bom_Pagador = Não | Tem_Casa_Própria = Não, Estado_Civil = Divorciado, Possui_Carro = Sim, Rendimentos = Alto) =

$$\frac{P(TCP=N \mid BP=N)*P(EC=D \mid BP=N)*P(PC=S \mid BP=N)*P(R=A \mid BP=N)*P(BP=N)}{P(TCP=N \mid BP=N)*P(EC=D \mid BP=N)*P(PC=S \mid BP=N)*P(R=A \mid BP=N)*P(BP=N)+P(TCP=N \mid BP=S)*P(EC=D \mid BP=S)*P(PC=S \mid BP=S)*P(R=A \mid BP=S)*P(BP=S)}$$

Para realizar os cálculos acima, é preciso descobrir algumas probabilidades, consultando os dados presentes na tabela acima. As probabilidades estão a seguir:

- $P(BP = S) = 5/11$, $P(BP = N) = 6/11$
- $P(TCP = N \mid BP = S) = 3/5$, $P(TCP = N \mid BP = N) = 4/6$
- $P(EC = D \mid BP = S) = 2/5$, $P(EC = D \mid BP = N) = 1/6$
- $P(PC = S \mid BP = S) = 3/5$, $P(PC = S \mid BP = N) = 3/6$
- $P(R = A \mid BP = S) = 2/5$, $P(R = A \mid BP = N) = 1/6$
- $P(TCP = N) = 7/11$
- $P(EC = D) = 3/11$
- $P(PC = S) = 6/11$
- $P(R = A) = 3/11$

P(Bom_Pagador = Sim | Tem_Casa_Própria = Não, Estado_Civil = Divorciado, Possui_Carro = Sim, Rendimentos = Alto) =

$$\frac{3/5*2/5*3/5*2/5*5/11}{3/5*2/5*3/5*2/5*5/11+4/6*1/6*3/6*1/6*6/11} = 0.84$$

Agora, a probabilidade de não ser um bom pagador, dadas as mesmas condições:

$$P(TCP = N \mid BP = N) * P(EC = D \mid BP = N) * P(PC = S \mid BP = N) * P(R = A \mid BP = N) * P(BP = N)$$

$$P(TCP = N) * P(EC = D) * P(PC = S) * P(R = A)$$

$$\frac{4/6*1/6*3/6*1/6*6/11}{4/6*1/6*3/6*1/6*6/11+3/5*2/5*3/5*2/5*5/11} = 0.16$$

Desse modo, conforme as probabilidades calculadas, o registro 12 é um bom pagador, de acordo com o método de Naive-Bayes.

A seguir, os cálculos para o Registro 13:

- P(Bom_Pagador = Sim | Tem_Casa_Própria = Não, Estado_Civil = Divorciado, Possui_Carro = Sim, Rendimentos = Alto) =

$$\frac{P(TCP=S \mid BP=S)*P(EC=S \mid BP=S)*P(PC=N \mid BP=S)*P(R=M \mid BP=S)*P(BP=S)}{P(TCP=S \mid BP=S)*P(EC=S \mid BP=S)*P(PC=N \mid BP=S)*P(R=M \mid BP=S)*P(BP=S)+P(TCP=S \mid BP=N)*P(EC=S \mid BP=N)*P(PC=N \mid BP=N)*P(R=M \mid BP=N)*P(BP=N)}$$

- P (Bom_Pagador = Não | Tem_Casa_Própria = Não, Estado_Civil = Divorciado, Possui_Carro = Sim, Rendimentos = Alto)

$$\frac{P(TCP=S \mid BP=N)*P(EC=S \mid BP=N)*P(PC=N \mid BP=N)*P(R=M \mid BP=N)*P(BP=N)}{P(TCP=S \mid BP=N)*P(EC=S \mid BP=N)*P(PC=N \mid BP=N)*P(R=M \mid BP=N)*P(BP=N)+P(TCP=S \mid BP=S)*P(EC=S \mid BP=S)*P(PC=N \mid BP=S)*P(R=M \mid BP=S)*P(BP=S)}$$

Para realizar os cálculos acima, é preciso descobrir algumas probabilidades, consultando os dados presentes na tabela acima. As probabilidades estão a seguir:

- $P(BP = S) = 5/11$, $P(BP = N) = 6/11$
- $P(TCP = S \mid BP = S) = 2/5$, $P(TCP = S \mid BP = N) = 2/6$
- $P(EC = S \mid BP = S) = 3/5$, $P(EC = S \mid BP = N) = 1/6$
- $P(PC = N \mid BP = S) = 2/5$, $P(PC = N \mid BP = N) = 3/6$
- $P(R = M \mid BP = S) = 3/5$, $P(R = M \mid BP = N) = 2/6$
- $P(TCP = S) = 4/11$
- $P(EC = S) = 4/11$
- $P(PC = N) = 5/11$
- $P(R = M) = 5/11$

P(Bom_Pagador = Sim | Tem_Casa_Própria = Não, Estado_Civil = Divorciado, Possui_Carro = Sim, Rendimentos = Alto) =

$$\frac{2/5 * 3/5 * 2/5 * 3/5 * 5/11}{2/5 * 3/5 * 2/5 * 3/5 * 5/11 + 2/6 * 1/6 * 3/6 * 2/6 * 6/11} = 0.84$$

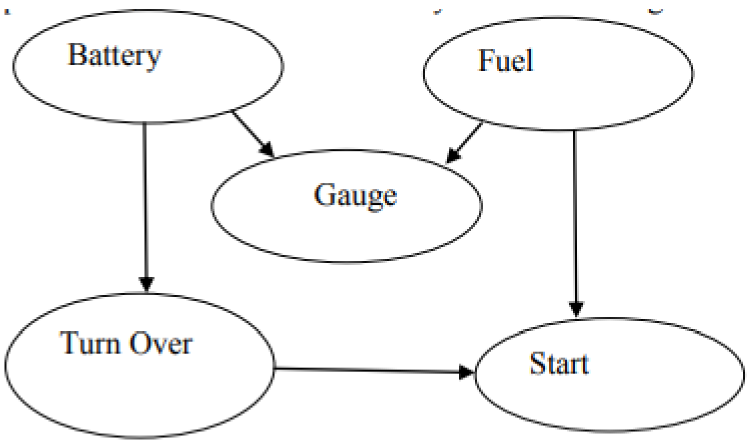
Agora, a probabilidade de não ser um bom pagador, dadas as mesmas condições:

P(Bom_Pagador = Não | Tem_Casa_Própria = Não, Estado_Civil = Divorciado, Possui_Carro = Sim, Rendimentos = Alto) =

$$\frac{2/6 * 1/6 * 3/6 * 2/6 * 6/11}{2/6 * 1/6 * 3/6 * 2/6 * 6/11 + 2/5 * 3/5 * 2/5 * 3/5 * 5/11} = 0.16$$

Desse modo, conforme as probabilidades calculadas, o registro 13 é um bom pagador, de acordo com o método de Naive-Bayes.

2) A rede bayesiana abaixo concerne ao problema de partida de um carro, de uma forma bem simplificada (Extraído do Livro Bayesian Reasoning and Machine Learning - D. Barber)



As variáveis aleatórias envolvidas são:
b=battery, g=gauge, f=fuel, t =turn over, s=start fa=false, tr= true
As probabilidades referentes a rede bayesiana são dadas por:

p(b=bad)=0.05	p(f=empty)=0.1
p(g=empty|b=good,f=not empty)=0.05	p(g=empty|b=good,f=empty)=0.98
p(g=empty|b=bad,f=not empty)=0.07	p(g=empty|b=bad,f=empty)=0.97
p(t=fa|b=good)=0.1	p(t=fa|b=bad)=0.96
p(s=fa|t=tr, f=not empty)=0.01	p(s=fa|t=tr, f=empty)=0.92
p(s=fa|t=fa, f=not empty)=1.0	p(s=fa|t=fa, f=empty)=0.99

Um agente inteligente com base nas inferências, isto é, no cálculo da P(f=empty|s=no) (a probabilidade do tanque estar vazio dado que o carro não deu partida) e da P(b=bad|s=no) (a probabilidade da bateria estar descarregada e o carro não deu partida), deve decidir qual o problema mais provável pela não partida do carro. Apresente a solução e implemente os cálculos de forma computacional.

Analisando o gráfico da rede bayesiana, pode tirar as seguintes conclusões:

- Battery (b) não é filho de ninguém;
- Fuel (f) não é filho de ninguém;
- Gauge (g) é filho de Battery (b) e Fuel (f);
- Turn Over (t) é filho de Battery (b);
- Start (s) é filho de Fuel (f).

De forma matemática, queremos calcular P(f=empty|s=fa). Assim, pela propriedade da independência condicional, que pela condição de Markov, podemos calcular a distribuição de probabilidade conjunta sobre todas as variáveis X1 , X2 , …, Xn da rede bayesiana através da seguinte fórmula:

$$P(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} P(x_i|Pais(x_i))$$

Para a questão, temos o seguinte cálculo:

$$P(f = empty|s = fa) = \frac{P(f = empty|s = fa)}{P(s = fa)} = \frac{P(f=empty)*P(s=fa|f=empty,t)}{P(s=fa)} \quad (2)$$

Calculando $P(s = fa|f = empty, t)$:

$$P(s = fa|f = empty, t) = P(s = fa|f = empty,\ t = tr) \cdot P(t = tr|b) + P(s = fa|f = empty,\ t = fa) \cdot P(t = fa|b) \quad (2.1)$$

É preciso, ainda, calcular $P(t = fa|b)$ e seu complementar $P(t = tr|b)$:

$$P(t = fa|b) = P(t = fa|b = bad) \cdot P(b = bad) + P(t = fa|b = good) \cdot P(b = good)$$
$$P(t = fa|b) = 0.96 * 0.05 + 0.1 * (1 - 0.05) = 0.143$$

Desse modo, $P(t = tr|b) = 1 - P(t = fa|b)$:

$$P(t = tr|b) = 1 - P(t = fa|b) = 1 - 0.143 = 0.857$$

Substituindo o resultado encontrado acima na equação 2.1:

$$P(s = fa|f = empty,\ t) = 0.92 * 0.857 + 0.\ 99 * 0.143 = 0.93001$$

Com os valores descobertos acima, ainda é preciso descobrir o valor de $P(s = fa)$, que será feito logo abaixo:

$$\begin{aligned} P(s = fa) = P(s = fa|f,\ t) = {} & P(s = fa|f = empty,\ t = tr) * P(t = tr|b) * P(f = empty) + \\ & P(s = fa|f = empty,\ t = fa) * P(t = fa|b) * P(f = empty) + \\ & P(s = fa|f = not\,empty,\ t = tr) * P(t = tr|b) * P(f = not\,empty) + \\ & P(s = fa|f = not\,empty,\ t = fa) * P(t = fa|b) * P(f = not\,empty) \end{aligned}$$

$$\begin{aligned} P(s = fa) = {} & 0.92 * 0.857 * 0.1 + 0.99 * 0.143 * 0.1 + \\ & 0.01 * 0.857 * (1 - 0.1) + 1.00 * 0.143 * (1 - 0.1) \end{aligned}$$

$$P(s = fa) = 0.2294$$

Com esses dados, podemos encontrar o valor para a equação 2:

$$P(f = empty|s = fa) = \frac{P(f=empty)*P(s=fa|f=empty,t)}{P(s=fa)} = \frac{0.1*0.93001}{0.2294} = 0.4054$$

Agora iremos calcular a outra probabilidade solicitada:

$$P(b = bad|s = fa) = \frac{P(b=bad, s=fa)}{P(s=fa)} = \frac{P(b=bad)\cdot P(s=fa|\, f, t)}{P(s=fa)} \quad (2.2)$$

É preciso descobrir $P(s = fa|\, f,\, t)$:

$$\begin{aligned} P(s = fa|\, f,\, t) = {} & P(s = fa|f = empty,\, t = tr)\cdot P(t = tr|b = bad)\cdot P(f = empty) + \\ & P(s = fa|f = empty,\, t = fa)\cdot P(t = fa|b = bad)\cdot P(f = empty) + \\ & P(s = fa|f = not\, empty,\, t = tr)\cdot P(t = tr|b = bad)\cdot P(f = not\, empty) + \\ & P(s = fa|f = not\, empty,\, t = fa)\cdot P(t = fa|b = bad)\cdot P(f = not\, empty) \end{aligned}$$

$$\begin{aligned} P(s = fa|\, f,\, t) = {} & 0.\,92\cdot 0.\,1\cdot(1 - 0.\,96) + 0.\,99\cdot 0.\,1\cdot 0.\,96 + \\ & 0.\,01\cdot 0.\,9\cdot(1 - 0.\,96) + 1\cdot 0.\,9\cdot 0.\,96 \end{aligned}$$

$$P(s = fa|\, f,\, t) = 0.\,96308$$

Substituindo os valores na equação 2.2:

$$P(b = bad|s = fa) = \frac{P(b=bad, s=fa)}{P(s=fa)} = \frac{P(b=bad)\cdot P(s=fa|\, f, t)}{P(s=fa)} = \frac{0.05*0.\,96308}{0.2294} = 0.\,2099$$

Como 0.4054 > 0.2099, a probabilidade do tanque está vazio dado que o carro não deu partida é maior.

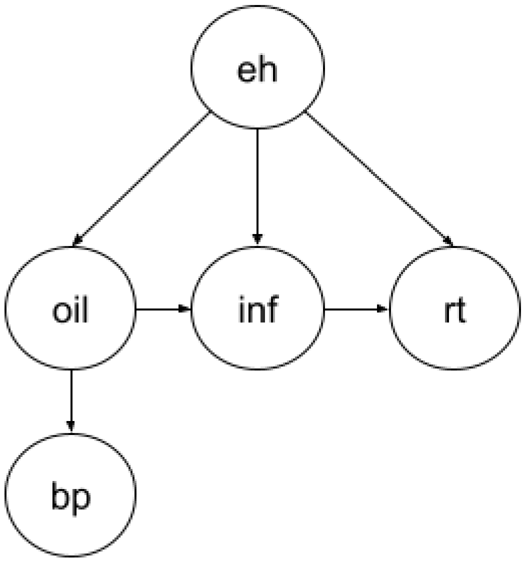
3) Uma rede de crença (ou rede bayesiana), modela a relação entre as variáveis: oil (price of oil), inf (inflation) , eh (economy health), bp (British Petroleum Stock price), rt (retailer stock price). Cada variável tem dois estados (l:low) e (h:high), exceto a variável bp que tem adicionalmente o estado (n: normal). A rede de crença modela as variáveis de acordo com a tabela abaixo. (Extraído do Livro Bayesian Reasoning and Machine Learning - D. Barber)

P(eh=l)=0.7	
P(bp=l|oil=l)=0.9	P(bp=n|oil=l)=0.1
P(bp=l|oil=h)=0.1	P(bp=n|oil=h)=0.4
P(oil=l|eh=l)=0.9	P(oil=l|eh=h)=0.05
P(rt=l|inf=l,eh=l)=0.9	P(rt=l|inf=l,eh=h)=0.1
P(rt=l|inf=h,eh=l)=0.1	P(rt=l|inf=h,eh=h)=0.01
P(inf=l|oil=l,eh=l)=0.9	P(inf=l|oil=l,eh=h)=0.1

P(inf=l|oil=h,eh=l)=0.1	P(inf=l|oil=h, eh=h)=0.01

a) Determine o gráfico da rede de crença (rede bayesiana) para este problema
b) Dado que a bp=n e rt=h, qual é a probabilidade de que a inflação seja alta?
Apresente a solução e implemente os cálculos de forma computacional

a)



b)

$$P(inf = h \mid bp = n, rt = h) = \frac{P(inf=h, bp=n, rt=h)}{P(bp=n, rt=h)}$$

$$P(inf = h \mid bp = n, rt = h) = \frac{P(inf=h \mid eh, oil) \cdot P(bp=n \mid oil) \cdot P(rt=h \mid eh, inf = h)}{P(bp=n, rt=h)}$$

(I):

$$P(inf = h \mid eh, oil) = P(inf = h \mid eh = l, oil = l) \cdot P(eh = l) \cdot P(oil = l \mid eh = l)$$
$$+ P(inf = h \mid eh = h, oil = l) \cdot P(eh = h) \cdot P(oil = l \mid eh = h)$$
$$+ P(inf = h \mid eh = l, oil = h) \cdot P(eh = l) \cdot P(oil = h)$$
$$+ P(inf = h \mid eh = h, oil = h) \cdot P(eh = h) \cdot P(oil = h \mid eh = h)$$

$$P(inf = h \mid eh, oil) = 0.1 \cdot 0.7 \cdot 0.9 + 0.9 \cdot 0.3 \cdot 0.05 + 0.9 \cdot 0.7 \cdot 0.1 + 0.99 \cdot 0.3 \cdot 0.95$$

$$P(inf = h \mid eh, oil) = 0.42165$$

(II):

$$P(bp = n \mid oil) = P(bp = n \mid oil = h) \cdot P(oil = h \mid eh) + P(bp = n \mid oil = l) \cdot P(bp = l \mid eh)$$
$$P(bp = n \mid oil = h) \cdot [P(oil = h \mid eh = l) \cdot P(eh = l) + P(oil = h \mid eh = h) \cdot P(eh = h)]$$
$$+ P(bp = n \mid oil = l) \cdot [P(oil = l \mid eh = l) \cdot P(eh = l) + P(oil = l \mid eh = h) \cdot P(eh = h)]$$

$P(bp = n \mid oil) = 0.4 \cdot [0.1 \cdot 0.7 + 0.95 \cdot 0.3] + 0.1 \cdot [0.9 \cdot 0.7 + 0.05 \cdot 0.3]$

$P(bp = n \mid oil) = 0.2065$

(III):

$P(rt = h \mid inf = h, eh) = P(rt = h \mid inf = h, eh = l) \cdot P(inf = h \mid eh = l, oil)^{(*)} \cdot P(eh = l)$
$+ P(rt = h \mid inf = h, eh = h) \cdot P(inf = h \mid eh = h, oil)^{(**)} \cdot P(eh = h) =$

$(\star)$: $P(inf = h \mid eh = l, oil) = P(inf = h \mid eh = l, oil = l) \cdot P(oil = l \mid eh = l) \cdot P(eh = l)$
$+ P(inf = h \mid eh = l, oil = h) \cdot P(oil = h \mid eh = l) \cdot P(eh = l) =$

$P(inf = h \mid eh = l, oil) = 0.1 \cdot 0.9 \cdot 0.7 + 0.9 \cdot 0.1 \cdot 0.7 = 0.126$

$(\star\star)$: $P(inf = h \mid eh = h, oil) = P(inf = h \mid eh = h, oil = l) \cdot P(oil = l \mid eh = h)$
$\cdot P(eh = h) + P(inf = h \mid eh = h, oil = h) \cdot P(eh = h) \cdot P(oil = h \mid eh = h) =$

$P(inf = h \mid eh = l, oil) = 0.99 \cdot 0.3 \cdot 0.95 + 0.9 \cdot 0.3 \cdot 0.05$

$P(inf = h \mid eh = l, oil) = 0.29565$

Voltando para P(rt=h | inf = h, eh), temos:

$P(rt = h \mid inf = h, eh) = 0.9 \cdot 0.126 \cdot 0.7 + 0.99 \cdot 0.29565 \cdot 0.3 = 0.1672$

Calculando P(bp = n, rt = h)

$P(bp = n \mid oil) = P(bp = n \mid oil = l) \cdot P(oil = l \mid eh) + P(bp = n \mid oil = h) \cdot$
$P(oil = h \mid eh)$

$P(bp = n \mid oil) = 0.2065$

$P(rt = h \mid inf, eh) = 0.1 \cdot 0.7 \cdot (1 - 0.126) + 0.9 \cdot 0.3 \cdot (1 - 0.29565) +$
$0.9 \cdot 0.7 \cdot 0.126 + 0.99 \cdot 0.3 \cdot 0.29565$

$P(rt = h \mid inf, eh) = 0.4185$

Com todos os valores, podemos substituir na equação principal:

$$P(inf = h \mid bp = n, rt = h) = \frac{P(inf=h \mid eh, oil) \cdot P(bp=n \mid oil) \cdot P(rt=h \mid eh, inf = h)}{P(bp=n, rt= h)}$$

$$P(inf = h \mid bp = n, rt = h) = \frac{0.42165 \cdot 0.2065 \cdot 0.1672}{0.2065 \cdot 0.4185} = 0.1685$$

4) Considere o problema de decisão caracterizado por uma sequência de eventos que podem ser apresentados por um gráfico conhecido como rede de decisão. Uma casa está a venda. A casa foi

construída a mais de dez anos. João está interessado em comprar a casa como investimento. Isto é fazer uma pequena reforma e revender a casa. Ele considera que a casa tem 70% de chance de estar realmente em bom estado. Se a casa estiver realmente em bom estado ele pode após uma pequena reforma ter um lucro de 30.000 reais na revenda. Caso contrário ele vai ter um prejuízo de 18.000,00 reais. João sabe que se ele contratar um profissional especializado em inspecionar imóveis ele terá uma melhor avaliação da situação da casa. Entretanto a contratação deste profissional requer um custo de 3.600,00 reais. A tabela abaixo indica as probabilidades envolvidas no processo de fazer ou não a inspeção e as condições do imóvel.

Tabela de Probabilidades

I	C	P(A=Bom|I,C)	P(A=Ruim|I,C)	P(A=N.I|I,C)
Sim	Bom	0.95	0.05	0
Sim	Não	0.10	0.95	0
Não	Bom	0	0	1
Não	Não	0	0	1

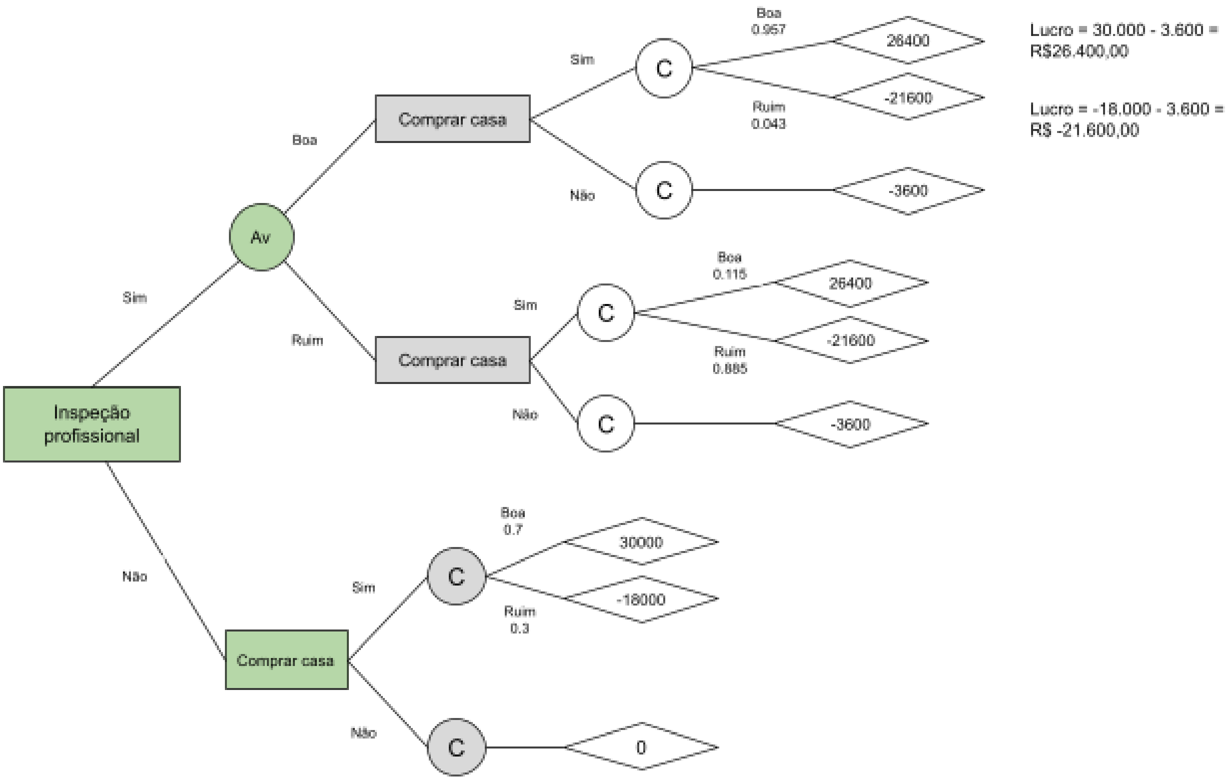
I: Inspecionar, C: Condições do imóvel, A: Avaliação

Tabela de Utilidades:

CC (Comprar Casa)	C (Condição)	U(CC, C)
Sim	Bom	30.000,00
Não	Ruim	0
Não	Bom	0
Sim	Ruim	-18.000,00

I (Inspecionar)	U(I)
Sim	-3.600,00
Não	0.0

a) A rede de decisão é da forma:



Calculando as probabilidades de uma avaliação ser boa ou ruim, temos que:

$P(A = Boa) = P(Cond = Boa) \cdot P(A = Boa \,|\, Cond = Boa) + P(A = Boa \,|\, C = Ruim) \cdot P(C = Ruim)$

$P(A = Boa) = 0.7 \cdot 0.95 + 0.10 \cdot 0.30 = 0.695$

Para uma avaliação ruim, basta calcularmos:

$P(A = Ruim) = 1 - 0.695 = 0.305$

Dado que temos uma avaliação boa e que João está disposto a comprar a casa, podemos calcular a condição real da casa estando boa a partir da seguinte equação:

$P(CC = Sim \,|\, A = Boa \,|\, Cond = Boa) = \frac{P(A=Boa, Cond= Boa) \cdot P(Cond=Boa)}{P(A = Boa)} = \frac{0.95 \cdot 0.70}{0.695} = 0.9568$

$P(CC = Sim \,|\, A = Boa \,|\, Cond = Boa) = 0.9568 \approx 0.957$

Com isso, podemos analisar a probabilidade de que a condição real da casa seja ruim fazendo:

$P(CC = Sim \,|\, A = Boa \,|\, Cond = Ruim) = 1 - 0.9568 = 0.0431 \approx 0.043$

Agora, tendo uma avaliação ruim, podemos calcular a condição real da casa estando ruim a partir da seguinte equação:

$P(CC = Sim \mid A = Ruim \mid Cond = Ruim) = \frac{P(A = Ruim, Cond = Ruim) \cdot P(Cond=Ruim)}{P(A=Ruim)} =$

$P(CC = Sim \mid A = Ruim \mid Cond = Ruim) = \frac{0.90 \cdot 0.30}{0.305} = 0.885$

A probabilidade para que a condição real da casa esteja boa é dada, de forma semelhante à situação anterior, de acordo com a equação:

$P(CC = Sim \mid A = Ruim \mid Cond = Boa) = 1 - 0.885 = 0.115$

b) A melhor decisão é comprar a casa caso ele tenha uma boa avaliação do profissional.

5) Considere o problema de tomada de decisão caracterizado por uma sequência de eventos que podem ser apresentado por um grafo conhecido como rede de decisão. O problema em questão consiste das escolhas e das decisões por parte de uma empresa de petróleo. Uma determinada empresa petrolífera obteve a concessão para explorar uma certa região. Os estudos anteriores (testes preliminares) estimam a probabilidade de existir petróleo nessa região em 20 %. A companhia pode optar por um novo teste, que custa US$ 100.000,00, sendo que, se realmente existe petróleo, esse teste dirá com uma probabilidade de 0.85 que existe, e se realmente não existe, dirá com probabilidade 0.70 que não existe. Considerando que o custo de perfuração será de US$ 1.000.000,00 e que, se for encontrado petróleo, a companhia receberá US$ 20.000.000,00 pela produção. Considere, portanto os seguintes eventos e os seus complementos: (i) Evento T (a companhia faz o teste); (ii) Evento F (o teste é favorável à existência de petróleo; (iii) Evento P (a companhia perfura o poço); (iv) Evento E (existe petróleo).

a-) Construa a rede indicando os nós de decisões e os nós ao acaso (variáveis aleatórias). Considere as funções de utilidade, representadas por losangos, como sendo o lucro =receita - despesas, calculado em cada percurso da arvore.
b-) Determine em cada nó dos percursos da árvore a utilidade esperada.
c-)Usando o critério da utilidade máxima esperada, determine a melhor decisão.
d-) Qual o valor esperado do lucro da companhia se for tomada a melhor decisão?
e-) Apresente também a solução deste problema através de um programa computacional e simule diferentes situações alterando o valor das probabilidades.

Observações:
(i) O evento inicial da árvore é se companhia faz ou não faz o teste.
(ii) Para cada evento tem o seu complementar: Exemplo: T: Faz o teste, ⌐T: Não faz o teste
(iii) Para o cálculo da utilidade esperada determine antes as probabilidades condicionais a posteriori com base no teorema de Bayes.

a) Dadas as probabilidades iniciais e que NT é referente ao novo teste e P ao petróleo:

Probabilidades iniciais:

P(P = Sim) = 0.2
P(P = Não) = 0.8
P(NT = Sim | P = Sim) = 0.85
P(NT = Não | P = Sim) = 0.15
P(NT = Sim | P = Não) = 0.30
P(NT = Não | P = Não) = 0.70

Dessa forma, podemos calcular a possibilidade favorável de existir petróleo, dado que se iniciou um novo teste.

$P(NT = Sim) = P(NT = Sim | P = Sim) \cdot P(P = Sim) + P(NT = Sim | P = Não) \cdot P(P = Não) \rightarrow P(NT = Sim) = 0.85 \cdot 0.2 + 0.30 \cdot 0.8 = 0.41$

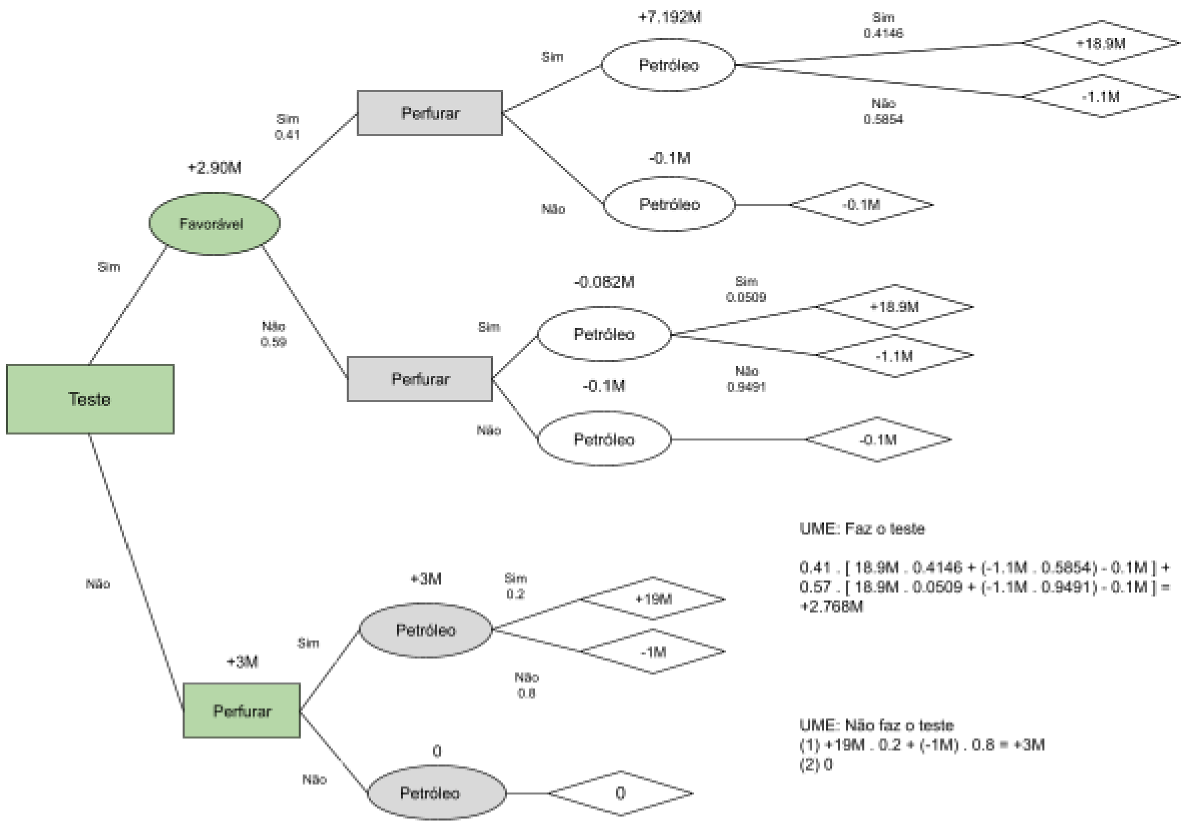
Assim, também podemos determinar P(NT = Não), como complemento à equação anterior:

$P(NT = Não) = 1 - 0.41 = 0.59$

Assim, temos que calcular as probabilidades de, caso haja um teste, de uma perfuração do solo, também temos que levar em consideração a possibilidade de achar ou não o petróleo.

$P(P = Sim | NT = Sim) = \frac{P(NT = Sim \mid P = Sim) \cdot P(P=Sim)}{P(NT = Sim)} = \frac{0.85 \cdot 0.2}{0.41} = 0.4146$
$P(P = Não | NT = Sim) = 1 - 0.4146 = 0.5853$
$P(P = Não | NT = Não) = \frac{P(NT = Não \mid P = Não)\, P(P = Não)}{P(NT = Não)} = \frac{0.70 \cdot 0.80}{0.59} = 0.9491$
$P(P = Sim | NT = Não) = 1 - 0.9491 = 0.0509$

Com todas as probabilidades em mãos, é possível montar a rede de decisão para analisarmos qual a melhor decisão a se tomar.



b) +7.192M, -0.1M, -0.082M, -0.1M, +2.9M, +3M, 0.
c) A partir do UME calculado, a melhor decisão é não fazer o teste, visto que o lucro é maior em relação quando se faz o teste.
d) O lucro esperado é de 3 milhões de dólares.

Trabalho 01 - Naive Bayes Anti-Spam

https://github.com/luizdts/artificial-intelligence/blob/main/Unidade%201/filtro_naivebayes.ipynb

Trabalho 02 - Aplicação do Random Forest

https://github.com/luizdts/artificial-intelligence/blob/main/Unidade%201/random-forest.ipynb

Referência para os trabalhos:
https://github.com/mohitgupta-omg/Kaggle-SMS-Spam-Collection-Dataset-/blob/master/Kaggle%20SMS%20Spam%20Collection.ipynb
https://www.kaggle.com/code/tcvieira/simple-random-forest-iris-dataset