CSE 250A: Assignment 2

Siyuan Liu, A53099540

October 12, 2015

2.1 Probabilistic inference

(a)

Some priors:

$$P(A, E, B) = P(A|E, B)P(E)P(B)$$

Calculating:

$$\begin{aligned}
P(E=1|A=1) &= \frac{P(E=1, A=1)}{P(A=1)} = \frac{\sum_b P(E=1, A=1, B=b)}{\sum_{e,b} P(A=1, E=e, B=b)} \\
&= \frac{0.95 \times 0.002 \times 0.001 + 0.29 \times 0.002 \times 0.999}{0.95 \times 0.002 \times 0.001 + 0.29 \times 0.002 \times 0.999 + 0.94 \times 0.998 \times 0.001 + 0.001 \times 0.998 \times 0.999} \\
&= 0.231008702
\end{aligned}$$

(b)

Calculating with same priors of (a):

$$\begin{aligned}
P(E=1|A=1, B=1) &= \frac{P(E=1, A=1, B=1)}{P(A=1, B=1)} = \frac{P(E=1, A=1, B=1)}{\sum_e P(A=1, E=e, B=1)} \\
&= \frac{0.95 \times 0.002 \times 0.001}{0.95 \times 0.002 \times 0.001 + 0.94 \times 0.998 \times 0.001} \\
&= 0.002021233591
\end{aligned}$$

(c)

Calculating:

$$\begin{aligned}
P(A=1|J=0) &= \frac{P(J=0|A=1)P(A=1)}{P(J=0)} \\
&= \frac{P(J=0|A=1)P(A=1)}{\sum_a P(J=0, A=a)} = \frac{P(J=0|A=1)P(A=1)}{\sum_a P(J=0|A=a)P(A=a)}
\end{aligned}$$

At this point:

$$\begin{aligned}
P(A=1) &= \sum_{e,b} P(A=1|E=e, B=b)P(E=e)P(B=b) \\
&= 0.002516442 \\
P(A=0) &= 1 - P(A=1) = 1 - 0.002516442 = 0.997483558
\end{aligned}$$

$$\begin{aligned}
P(A=1|J=0) &= \frac{P(J=0|A=1)P(A=1)}{\sum_a P(J=0|A=a)P(A=a)} \\
&= \frac{0.10 \times 0.002516442}{0.10 \times 0.002516442 + 0.95 \times 0.997483558} \\
&= 0.00026548638
\end{aligned}$$

(d)

Some priors:

$$P(A,J,M) = P(J,M|A)P(A) = P(J|A)P(M|A)P(A)$$

$$\begin{aligned}
P(A=1|J=0,M=0) &= \frac{P(J=0,M=0|A=1)P(A=1)}{P(J=0,M=0)} \\
&= \frac{P(J=0|A=1)P(M=0|A=1)P(A=1)}{\sum_a P(J=0,M=0,A=a)} \\
&= \frac{P(J=0|A=1)P(M=0|A=1)P(A=1)}{\sum_a P(J=0|A=a)P(M=0|A=a)P(A=a)} \\
&= \frac{0.10 \times 0.30 \times 0.002516442}{0.10 \times 0.30 \times 0.002516442 + 0.95 \times 0.99 \times 0.997483558} \\
&= 0.0000804653
\end{aligned}$$

(e)

$$\begin{aligned}
P(A=1|M=1) &= \frac{P(M=1|A=1)P(A=1)}{P(M=1)} \\
&= \frac{P(M=1|A=1)P(A=1)}{\sum_a P(M=1,A=a)} = \frac{P(M=1|A=1)P(A=1)}{\sum_a P(M=1|A=a)P(A=a)} \\
&= \frac{0.70 \times 0.002516442}{0.70 \times 0.002516442 + 0.01 \times 0.997483558} \\
&= 0.15009011775
\end{aligned}$$

(f)

$$\begin{aligned}
P(A=1|M=1,E=0) &= \frac{P(M=1,E=0|A=1)P(A=1)}{P(M=1,E=0)} \\
&= \frac{P(M=1|A=1)P(E=0|A=1)P(A=1)}{\sum_a P(M=1|A=a)P(E=0|A=a)P(A=a)} \\
&= \frac{0.70 \times (1-0.231008702) \times 0.002516442}{0.70 \times (1-0.231008702) \times 0.002516442 + 0.01 \times 0.99857774096 \times 0.997483558} \\
&= 0.11971342083
\end{aligned}$$

(commonsense patterns of reasoning) explaining away

2.2 Probabilistic reasoning

(a)

With Bayes Rule:

$$\begin{aligned}
r_k &= \frac{P(D=0|S_1=1,...,S_k=1)}{P(D=0|S_1=1,...,S_k=1)} \\
&= \frac{\frac{P(S_1=1,...,S_k=1|D=0)P(D=0)}{P(S_1=1,...,S_k=1)}}{\frac{P(S_1=1,...,S_k=1|D=1)P(D=1)}{P(S_1=1,...,S_k=1)}} \\
&= \frac{P(S_1=1,...,S_k=1|D=0)P(D=0)}{P(S_1=1,...,S_k=1|D=1)P(D=1)} \\
&= \frac{P(S_1=1|D=0)...P(S_k=1|D=0)P(D=0)}{P(S_1=1|D=1)...P(S_k=1|D=1)P(D=1)} \\
&= \frac{1 \times \frac{f(1)}{f(2)} \times \frac{f(2)}{f(3)} \times ... \times \frac{f(k-1)}{f(k)} \times \frac{1}{2}}{\frac{1}{2^k} \times \frac{1}{2}} \\
&= \frac{\frac{f(1)}{f(k)}}{\frac{1}{2^k}} = \frac{2^k}{2^k + (-1)^k}
\end{aligned}$$

So we can see, if k is even, $r_k < 1$, it means the doctor diagnoses the patient with the $D = 1$ form of the disease; if if k is odd, $r_k > 1$, it means the doctor diagnoses the patient with the $D = 0$ form of the disease.

(b)

The diagnosis does not become more or less certain as more symptoms are observed.

Because the diagnosis of form of disease D takes turns. It means if the result is $D = 0$ today, tomorrow it will be $D = 1$. As time goes by, given the method (Suppose that on the k^{th} day of the month, a test is done to determine whether the patient is exhibiting the k^{th} symptom) and diagonosis(each such test returns a positive result), the form of disease D are uniformly likely to be observed.

2.3 Sigmoid function

(a)

$$\begin{aligned}
\sigma'(z) &= -\frac{1}{(1+e^{-z})^2} e^{-z}(-1) \\
&= \frac{e^{-z}}{(1+e^{-z})^2} = \frac{1}{(1+e^{-z}) \times \frac{1+e^{-z}}{e^{-z}}} \\
&= \frac{1}{(1+e^{-z})(1+e^{z})} \\
&= \sigma(z)\sigma(-z)
\end{aligned}$$

(b)

$$\begin{aligned}
\sigma(z) + \sigma(-z) &= \frac{1}{1+e^{-z}} + \frac{1}{1+e^{z}} \\
&= \frac{(1+e^{z}) + (1+e^{-z})}{(1+e^{z})(1+e^{-z})} \\
&= \frac{2+e^{z}+e^{-z}}{2+e^{z}+e^{-z}} \\
&= 1
\end{aligned}$$

(c)

$$\begin{aligned}
L(\sigma(z)) &= log(\frac{\sigma(z)}{1-\sigma(z)}) \\
&= log(\frac{\frac{1}{1+e^{-z}}}{1-\frac{1}{1+e^{-z}}}) \\
&= log(\frac{1}{1+e^{-z}}) - log(1-\frac{1}{1+e^{-z}}) \\
&= -log(1+e^{-z}) - (log(e^{-z}) - log(1+e^{-z})) = -log(e^{-z}) \\
&= z
\end{aligned}$$

2.4 Conditional independence

1. {sprinkler, rain, {month}}
2. {month, water, {sprinkler, rain}}
3. {month, water, {sprinkler, rain, fall}}
4. {month, fall, {water}}
5. {month, fall, {sprinkler, rain}}
6. {month, fall, {water, sprinkler}}
7. {month, fall, {water, rain}}
8. {month, fall, {water, sprinkler, rain}}
9. {sprinkler, fall, {water}}
10. {sprinkler, fall, {water, month}}
11. {sprinkler, fall, {water, rain}}
12. {sprinkler, fall, {water, month, rain}}
13. {rain, fall, {water}}
14. {rain, fall, {water, month}}
15. {rain, fall, {water, sprinkler}}
16. {rain, fall, {water, month, sprinkler}}

2.5 Markov blanket

A Markov blanket is shown in Figure 1, the nodes contains 1)X, 2)its parents, 3)its children, 4)parents of its children.

Look at Figure 2, nodes outside the Markov blanket of X can only contains 1)parents of its parents(Node 1), 2)parents of parents of its children(Node 2), 3)children of its parents(Node 3), 4)children of parents of its children(Node 4), 5)children of its children(Node 5).

Truly, there are more nodes existing out of the blanket. If all these nodes (Node 1, 2, 3, 4, 5) are conditional independent with X given B_X, all the paths from Node 1, 2, 3, 4, 5 to X are d-separated. Therefore, all the paths from more nodes out of the blanket to X are just extention of paths from Node 1, 2, 3, 4, 5 to X. So these paths are also d-separated, which means they are also conditional independent with X given $B-X$. In other words, we just need to consider these five kinds of nodes.

Because the network can be seen as a kind of tree, there is just one path from a node to source node X:
Paths($1->U_1->X$) from node 1 to X is d-separated for consition I
Paths($2->Z_1->Y_1-X$) from node 2 to X is d-separated for consition I
Paths($3<-U_2->X$) from node 3 to X is d-separated for consition II
Paths($4<-Z_1->Y_1-X$) from node 4 to X is d-separated for consition II
Paths($5<-Y_2<-X$) from node 5 to X is d-separated for consition I

To sum up, any node Y outside the Markov blanket of X (that is, satisfying $Y \notin B_X$ and $Y \notin X$), we have:

$$P(X, Y|B_X) = P(X|B_X)P(Y|B_X)$$

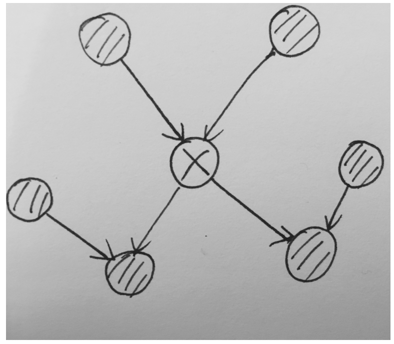

Figure 1: Original Markov Blanket

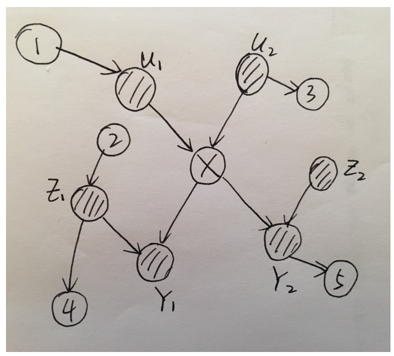

Figure 2: Original Markov Blanket with Outside Nodes

2.6 Noisy-OR

$$\begin{aligned}
P(Z = 1|X = 0, Y = 0) &< P(Z = 1|X = 0, Y = 1) \\
P(Z = 1|X = 1, Y = 0) &< P(Z = 1|X = 0, Y = 1) \\
P(Z = 1|X = 1, Y = 0) &< P(Z = 1|X = 1, Y = 1) \\
P(X = 1) &> P(X = 1|Z = 1) \\
P(X = 1) &= P(X = 1|Y = 1) \\
P(X = 1|Z = 1) &> P(X = 1|Y = 1, Z = 1) \\
P(X = 1)P(Y = 1)P(Z = 1) &< P(X = 1, Y = 1, Z = 1)
\end{aligned}$$

2.7 More conditional independence

false $P(E, F|D) = P(E|D)P(F|D)$
false $P(E, F|C, D) = P(E|C, D)P(F|C, D)$
true $P(E, F|A, B, D) = P(E|A, B, D)P(F|A, B, D)$
false $P(D|C) = P(D)$
true $P(D|A, B) = P(D|A, B, C)$
true $P(A, B) = P(A)P(B)$
false $P(A|C, D) = P(A|C, D, F)$
true $P(A|B, C, D) = P(A|B, C, D, F)$

true $\quad P(B|A,C,D,F) = P(B|A,C,D,F,E)$
false $\quad P(B,F,A,E|C,D) = P(B,F|C,D)P(A,E|C,D)$

2.8 Even more conditional independence

S = {A, C, E, D, G}
S = {D, G, F}
S = {B, D, G}
S = {F, G}
S = {A, E, F}
S = {F}
S = {A, B, C, D, F, G}
S = {$\emptyset$}
S = {C, D, E, G}
S = {$\emptyset$}