CS688: Graphical Models - Spring 2016

Assignment 1 Report

Shamya Karumbaiah

1. By factorization, the Bayesian network joint distribution for this graph (say BN) is given by -
$P(\textbf{BN}) = P(G, A, BP, CH, HD, CP, EIA, ECG, HR)$
$= P(G)P(A)P(BP|G)P(CH|G, A)P(HD|BP, CH)P(CP|HD)P(EIA|HD)P(ECG|HD)P(HR|HD, A)$

An alternate way to derive this is to apply chain rule at the joint distribution as below-
$P(\textbf{BN}) = P(G, A, BP, CH, HD, CP, EIA, ECG, HR)$
$= P(G)P(A|G)P(BP|G, A)P(CH|G, A, BP)P(HD|G, A, BP, CH)P(CP|G, A, BP, CH, HD)P(EIA|G, A, BP, CH, HD, CP)P(ECG|G, A, BP, CH, HD, CP, EIA)P(HR|G, A, BP, CH, HD, CP, EIA, ECG)$

And now use the conditional independence from the local markov property to simplify this equation-
$= P(G)P(A)P(BP|G)P(CH|G, A)P(HD|BP, CH)P(CP|HD)P(EIA|HD)P(ECG|HD)P(HR|HD, A)$

2. The log likelihood of a Bayesian Network as a function of the parameters θ given N data cases is given by -
$\mathcal{L}(\theta|\textbf{x}_{1:N}) = \frac{1}{N}\sum_{n=1}^{N}\sum_{d=1}^{D}\sum_{v=1}^{V_d}[x_{nd} = v]log\theta^{X_d}_{v|\textbf{x}_n, Pa(X_d)}$

Expanding this for the given bayesian network, we get the log likelihood as follows -
$\mathcal{L}(\theta|\textbf{x}_{1:N}) = \frac{1}{N}\sum_{n=1}^{N}(\sum_{g\in Val(G)}[g_n = g]log\theta^G_g + \sum_{a\in Val(A)}[a_n = a]log\theta^A_a + \sum_{bp\in Val(BP)}[bp_n = bp]log\theta^{BP}_{bp|g_n} + \sum_{ch\in Val(CH)}[ch_n = ch]log\theta^{CH}_{ch|g_n,a_n} + \sum_{hd\in Val(HD)}[hd_n = hd]log\theta^{HD}_{hd|bp_n,ch_n} + \sum_{cp\in Val(CP)}[cp_n = cp]log\theta^{CP}_{cp|hd_n} + \sum_{eia\in Val(EIA)}[eia_n = eia]log\theta^{EIA}_{eia|hd_n} + \sum_{ecg\in Val(ECG)}[ecg_n = ecg]log\theta^{ECG}_{ecg|hd_n} + \sum_{hr\in Val(HR)}[hr_n = hr]log\theta^{HR}_{hr|hd_n,a_n})$

3. The maximum likelihood estimate for $\theta^{HR}_{hr|a,hd}$ could be derived from below optimization problem -

$$\arg\max_{\theta^{HR}_{.|a,hd}} \frac{1}{N}\sum_{n=1}^{N}[a_n = a, hd_n = hd]\sum_{hr\in Val(HR)}[hr_n = hr]log\theta^{HR}_{hr|a,hd}$$

subject to $\sum_{hr\in Val(HR)}\theta^{HR}_{hr|a,hd} = 1$

since $Val(HR) \in \{L, H\}$, our likelihood function can be written as below
$\mathcal{L}(\theta^{HR}_{.|1,Y}\textbf{x}_{1:N}) = \frac{1}{N}\sum_{n=1}^{N}([a_n = 1, hd_n = Y, hr_n = L]log\theta^{HR}_{L|1,Y} + [a_n = 1, hd_n = Y, hr_n = H]log\theta^{HR}_{H|1,Y})$

since $\theta^{HR}_{H|1,Y} = 1 - \theta^{HR}_{L|1,Y}$, we can rewrite the above equation as
$\mathcal{L}(\theta^{HR}_{.|1,Y}\textbf{x}_{1:N}) = \frac{1}{N}\sum_{n=1}^{N}([a_n = 1, hd_n = Y, hr_n = L]log\theta^{HR}_{L|1,Y} + [a_n = 1, hd_n = Y, hr_n = H]log(1 - \theta^{HR}_{L|1,Y}))$

To find the maxima, lets differentiate this equation and set it to 0,
$\frac{1}{N}(\sum_{n=1}^{N} \frac{[a_n = 1, hd_n = Y, hr_n = L]}{\theta^{HR}_{L|1,Y}} + \sum_{n=1}^{N} \frac{[a_n = 1, hd_n = Y, hr_n = H]}{1 - \theta^{HR}_{L|1,Y}}) = 0$

Let $N_{L,1,Y} = \sum_{n=1}^{N}[hr_n = L, a_n = 1, hd_n = Y]$ and $N_{H,1,Y} = \sum_{n=1}^{N}[hr_n = H, a_n = 1, hd_n = Y]$. We have $N_{1,Y} = N_{L,1,Y} + N_{H,1,Y}$

Substituting this in the previous equation, we get
$\frac{1}{N}(\frac{N_{L,1,Y}}{\theta^{HR}_{L|1,Y}} + \frac{N_{1,Y} - N_{L,1,Y}}{1 - \theta^{HR}_{L|1,Y}}) = 0$

Solving this, we get
$\theta^{HR}_{L|1,Y} = \frac{N_{L,1,Y}}{N_{1,Y}}$

To verify if this maxima is the global maxima, lets take the second derivative
$\frac{\partial}{\partial\theta}(\frac{1}{N}(\frac{N_{L,1,Y}}{\theta^{HR}_{L|1,Y}} - \frac{N_{H,1,Y}}{1 - \theta^{HR}_{L|1,Y}}))$
$= \frac{1}{N}(-\frac{N_{L,1,Y}}{(\theta^{HR}_{L|1,Y})^2} - \frac{N_{H,1,Y}}{(1 - \theta^{HR}_{L|1,Y})^2})$

This is always a negative value if $N_{1,2} > 0$ Hence, this is a global maxima and thus the maximum likelihood estimate for $\theta^{HR}_{L|1,Y} = \frac{N_{L,1,Y}}{N_{1,Y}}$

4. The CPTs are listed below -

(a) $P_\theta(A)$

$P_\theta(A)$	A
0.1770	<45
0.3086	45-55
0.5144	>=55

(b) $P_\theta(BP|G)$

$P_\theta(BP\|G)$	BP	G
0.3659	Low	Female
0.6341	High	Female
0.4720	Low	Male
0.5280	High	Male

(c) $P_\theta(HD|BP, CH)$

$P\theta(HD \mid BP, CH)$	HD	BP	CH
0.5263	N	Low	Low
0.4737	Y	Low	Low
0.5909	N	High	Low
0.4091	Y	High	Low
0.5862	N	Low	High
0.4138	Y	Low	High
0.5130	N	High	High
0.4870	Y	High	High

(d) $P_\theta(HR|A, HD)$

$P\theta(HR \mid A, HD)$	HR	A	HD
0.0606	L	<45	N
0.9394	H	<45	N
0.1731	L	45-55	N
0.8269	H	45-55	N
0.3333	L	>=55	N
0.6667	H	>=55	N
0.6000	L	<45	Y
0.4000	H	<45	Y
0.5217	L	45-55	Y
0.4783	H	45-55	Y
0.5714	L	>=55	Y
0.4286	H	>=55	Y

Code - **Question4Code.m** gives the code to calculate the above CPTs for the given random variables in this bayesian network.

5. (a) $P(CH|A = 2, G = M, CP = None, BP = L, ECG = Normal, HR = L, EIA = No, HD = No)$

Since the values of parents of CH is given ($Pa(CH) \in (G, A)$), we could use the local markov property and say, $CH \perp BP|G, A$. Hence,

$P(CH|A = 2, G = M, CP = None, BP = L, ECG = Normal, HR = L, EIA = No, HD = No)$
= $P(CH|A = 2, G = M, CP = None, ECG = Normal, HR = L, EIA = No, HD = No)$

Lets consider the case of $CH = Low$. Using Bayes rule, we can write this as

$P(CH = Low|A = 2, G = M, CP = None, ECG = Normal, HR = L, EIA = No, HD = No)$

$$= \frac{P(CH = Low, A = 2, G = M, CP = None, ECG = Normal, HR = L, EIA = No, HD = No)}{P(A = 2, G = M, CP = None, ECG = Normal, HR = L, EIA = No, HD = No)}$$

$$= \frac{P(CH = Low, A = 2, G = M, CP = None, ECG = Normal, HR = L, EIA = No, HD = No)}{\sum_{ch \in Val(CH)} P(CH = ch, A = 2, G = M, CP = None, ECG = Normal, HR = L, EIA = No, HD = No)}$$

Expressing this in the factorized joint distribution for the Bayesian network,

$$=\frac{\begin{array}{r}P(G=M)P(A=2)P(CH=Low|G=M,A=2)P(HD=No|BP=L,CH=Low)\\ P(CP=None|HD=No)P(EIA=No|HD=No)\\ P(ECG=Normal|HD=No)P(HR=L|HD=No,A=2)\end{array}}{\begin{array}{r}\sum_{ch\in Val(CH)}(P(G=M)P(A=2)P(CH=ch|G=M,A=2)\\ P(HD=No|BP=L,CH=ch)P(CP=None|HD=No)P(EIA=No|HD=No)\\ P(ECG=Normal|HD=No)P(HR=L|HD=No,A=2)\end{array}}$$

Since, value of all the other nodes is known, this can be simplified to -

$$=\frac{P(CH=Low|G=M,A=2)P(HD=No|BP=L,CH=Low)}{\sum_{ch\in Val(CH)}P(CH=ch|G=M,A=2)P(HD=No|BP=L,CH=ch)}$$

$$=\frac{0.1667X0.5263}{(0.1667X0.5263)+(0.8333X0.5862)}$$

$=$ **0.1522596**

$P(CH=High|A=2,G=M,CP=None,BP=L,ECG=Normal,HR=L,EIA=No,HD=No)$
$=1-P(CH=Low|A=2,G=M,CP=None,BP=L,ECG=Normal,HR=L,EIA=No,HD=No)$
$=$ **0.8477404**

Below is the table giving the distribution over CH -

P(CH=Low)	0.1522596
P(CH=High)	0.8477404

(b) $P(BP|A=2,CP=Typical,CH=H,ECG=Normal,HR=H,EIA=Yes,HD=No)$

Since, the value of the random variable G ($Pa(BP)$) is not given, the conditional independence $CH\perp BP|G$ **doesn't** hold. Hence, we must consider all the random variables including BP.

Lets consider the case of $BP=Low$. Using Bayes rule, we can write this as

$$P(BP=Low|A=2,CP=Typical,CH=H,ECG=Normal,HR=H,EIA=Yes,HD=No)$$

$$=\frac{\begin{array}{r}P(BP=Low,A=2,CP=Typical,CH=H,ECG=Normal\\ ,HR=H,EIA=Yes,HD=No)\end{array}}{P(A=2,CP=Typical,CH=H,ECG=Normal,HR=H,EIA=Yes,HD=No)}$$

$$= \frac{\sum_{g \in Val(G)} P(BP = Low, A = 2, G = g, CP = Typical, CH = H, ECG = Normal, HR = H, EIA = Yes, HD = No)}{\sum_{g \in Val(G)} \sum_{bp \in Val(BP)} P(BP = bp, A = 2, G = g, CP = Typical, CH = H, ECG = Normal, HR = H, EIA = Yes, HD = No)}$$

Expressing this in the factorized joint distribution for the Bayesian network,

$$= \frac{\sum_{g \in Val(G)} P(G = g)P(A = 2)P(BP = Low|G = g)P(CH = H|G = g, A = 2) P(HD = No|BP = Low, CH = H)P(CP = Typical|HD = No) P(EIA = Yes|HD = No)P(ECG = Normal|HD = No)P(HR = H|HD = No, A = 2)}{\sum_{g \in Val(G)} \sum_{bp \in Val(BP)} P(G = g)P(A = 2)P(BP = bp|G = g) P(CH = H|G = g, A = 2)P(HD = No|BP = bp, CH = H)P(CP = Typical|HD = No) P(EIA = Yes|HD = No)P(ECG = Normal|HD = No)P(HR = H|HD = No, A = 2)}$$

Since, value of most other nodes is known, this can be simplified to -

$$= \frac{\sum_{g \in Val(G)} P(G = g)P(BP = Low|G = g)P(CH = H|G = g, A = 2)P(HD = No|BP = Low, CH = H)}{\sum_{g \in Val(G)} \sum_{bp \in Val(BP)} P(G = g)P(BP = bp|G = g)P(CH = H|G = g, A = 2)P(HD = No|BP = bp, CH = H)}$$

Substituting the values we get,
$=$ **0.451336**

$P(BP = High|A = 2, CP = Typical, CH = H, ECG = Normal, HR = H, EIA = Yes, HD = No)$
$= 1 - P(BP = Low|A = 2, CP = Typical, CH = H, ECG = Normal, HR = H, EIA = Yes, HD = No)$
$=$ **0.548664**

Below is the table giving the distribution over BP -

P(BP=Low)	0.451336
P(BP=High)	0.548664

Code - **Question5Code.m** gives the code to calculate the CPTs for **all** the random variables in this bayesian network. This is used to pick the necessary values to substitute in the above equations. The maximum likelihood estimate for all the parameters are listed in these CPTs too.

6. (b) Using the conditional independence property, we have $HD \perp G|BP, CH$ Hence,

$P(HD|G, A, BP, CH, CP, EIA, ECG, HR)$
$= P(HD|A, BP, CH, CP, EIA, ECG, HR)$

$$= \frac{P(HD, A, BP, CH, CP, EIA, ECG, HR)}{P(A, BP, CH, CP, EIA, ECG, HR)}$$

$$= \frac{P(HD, A, BP, CH, CP, EIA, ECG, HR)}{\sum_{hd \in Val(HD)} P(HD, A, BP, CH, CP, EIA, ECG, HR)}$$
$$= \frac{P(A)P(BP|G)P(CH|G, A)P(HD|BP, CH)P(CP|HD)P(EIA|HD)P(ECG|HD)P(HR|HD, A)}{\sum_{hd \in Val(HD)} P(A)P(BP|G)P(CH|G, A)P(HD|BP, CH)P(CP|HD)P(EIA|HD)P(ECG|HD)P(HR|HD, A)}$$

Canceling the common terms, we have -

$$= \frac{P(HD|BP, CH)P(CP|HD)P(EIA|HD)P(ECG|HD)P(HR|HD, A)}{\sum_{hd \in Val(HD)} P(HD|BP, CH)P(CP|HD)P(EIA|HD)P(ECG|HD)P(HR|HD, A)}$$

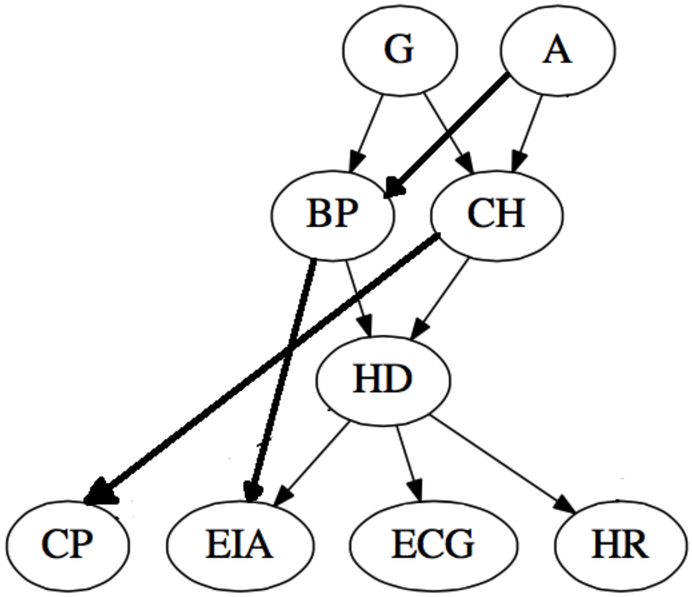
(c) Below is the table giving the result of the test run -

FileID	Accuracy
1	0.73
2	0.80
3	0.67
4	0.80
5	0.78

The mean prediction accuracy over the five test files is **0.756** or 75.6% and the standard deviation is **0.05595** and the variance is 0.002504.

Code - **Question6Code.m** gives the code to calculate the CPTs for all the random variables in the given bayesian network. The maximum likelihood estimate for all the parameters are listed in these CPTs too. The code trains on the train data in file *i* and tests in on the test data in the corresponding file *i*. It also gives the classification accuracy and the count of misclassified data points.

7: (a) The graphical model for the network is as below -



(b) By factorization, the Bayesian network joint distribution for this graph (say BN1) is given by -
$P(\textbf{BN1}) = P(G, A, BP, CH, HD, CP, EIA, ECG, HR) =$
$P(G)P(A)P(BP|G, A)P(CH|G, A)P(HD|BP, CH)P(CP|CH)P(EIA|HD, BP)P(ECG|HD)P(HR|HD)$

(c) Age contributes directly to BP too. But this additional edge doesn't have a direct impact on the heart disease classification. Also, age may not have a direct impact on HR. EIA could be correlated to BP and CH could be correlated to EIA. The dependency between HD and CP is also removed. Changing this made the mean classification accuracy of heart disease to improve by 2.6% (results in the next section).

(d) Below is the table giving the result of the test run -

FileID	Accuracy
1	0.78
2	0.83
3	0.70
4	0.80
5	0.80

The mean prediction accuracy over the five test files is **0.782** or 78.2% and the standard deviation is **0.04919** and the variance is 0.00412. The mean classification accuracy of heart disease improved by 2.6%.

Code - **Question7Code.m** gives the code to calculate the CPTs for all the random variables in this bayesian network. The maximum likelihood estimate for all the parameters are listed in these CPTs too. The code trains on the train data in file *i* and tests in on the test data in the corresponding file *i*. It also gives the classification accuracy and the count of misclassified data points.