Homework 3

Chen Lu

June 29, 2018

1 Probabilistic Graphical Models

1.1 Conditional Queries in a Bayesian Network

Answer
1. the Markov blanket of G3 are G1 and X3.
2. Initial factors:$p(G_1)p(G_2|G_1)p(X_1|G_1)p(G_3|G_1)p(X_3|G_3)p(X_2|G_2)$
The elimination order: X_3, X_1, G_3, G_2
Step 1: conditioning

$$m_{X_2}(G_2) = p(X_2 = 50|G_2) = \sum_{X_2} p(X_2|G_2)\delta(X_2 = 50)$$

Step 2: Eliminate X_3, X_1, G_3, G_2 in order.

$$\begin{aligned}
&\sum_{X_3}\sum_{G_3}\sum_{X_1}\sum_{G_2} p(G_1)p(G_2|G_1)p(X_1|G_1)p(G_3|G_1)p(X_3|G_3)p(X_2|G_2)\\
&= \sum_{X_3}\sum_{G_3}\sum_{X_1}\sum_{G_2} p(G_1)p(G_2|G_1)p(X_1|G_1)p(G_3|G_1)p(X_3|G_3)m_{X_2}(G_2)\\
&= \sum_{G_3}\sum_{X_1}\sum_{G_2} p(G_1)p(G_2|G_1)p(X_1|G_1)p(G_3|G_1)m_{X_2}(G_2)\\
&= \sum_{G_3}\sum_{G_2} p(G_1)p(G_2|G_1)p(G_3|G_1)m_{X_2}(G_2)\\
&= \sum_{G_2} p(G_1)p(G_2|G_1)m_{X_2}(G_2)\\
&= p(G_1)m_{G_2}(G_1)
\end{aligned}$$

$$
\begin{aligned}
p(G_1|X_2=50) &= \frac{p(G_1, X_2=50)}{p(X_2=50)} \\
&= \frac{p(G_1)m_{G_2}(G_1)}{\sum_{G_1} p(G_1)m_{G_2}(G_1)} \\
&= \frac{m_{G_2}(G_1)}{\sum_{G_1} m_{G_2}(G_1)}
\end{aligned}
$$

$$
\begin{aligned}
m_{G_2}(G_1=1) &= \sum_{G_2} p(G_2|G_1=1)m_{X_2}(G_2) \\
&= p(G_2=1|G_1=1)p(X_2=50|G_2=1) + p(G_2=2|G_1=1)p(X_2=50|G_2=1) \\
&= 0.9N(X_2=50|\mu=55, \sigma^2=10) + 0.1N(X_2=50|\mu=65, \sigma^2=10)
\end{aligned}
$$

$$
\begin{aligned}
m_{G_2}(G_1=2) &= \sum_{G_2} p(G_2|G_1=2)m_{X_2}(G_2) \\
&= p(G_2=1|G_1=2)p(X_2=50|G_2=1) + p(G_2=2|G_1=2)p(X_2=50|G_2=1) \\
&= 0.1N(X_2=50|\mu=55, \sigma^2=10) + 0.9N(X_2=50|\mu=65, \sigma^2=10)
\end{aligned}
$$

Thus

$$
\begin{aligned}
p(G_1=1|X_2=50) &= \frac{0.9N(X_2=50|\mu=55, \sigma^2=10) + 0.1N(X_2=50|\mu=65, \sigma^2=10)}{N(X_2=50|\mu=55, \sigma^2=10) + N(X_2=50|\mu=65, \sigma^2=10)} = 0.9 \\
p(G_1=2|X_2=50) &= \frac{0.1N(X_2=50|\mu=55, \sigma^2=10) + 0.9N(X_2=50|\mu=65, \sigma^2=10)}{N(X_2=50|\mu=55, \sigma^2=10) + N(X_2=50|\mu=65, \sigma^2=10)} = 0.1
\end{aligned}
$$

3. Step 1: Conditioning

$$
m_{X_2}(G_2) = p(X_2=50|G_2) = \sum_{X_2} p(X_2|G_2)\delta(X_2=50)
$$

$$
m_{X_3}(G_3) = p(X_3=50|G_3) = \sum_{X_3} p(X_3|G_3)\delta(X_3=50)
$$

Step 2: Eliminate X_1

$$\begin{aligned}
p(G_1, X_2 = 50, X_3 = 50) &= \sum_{X_1}\sum_{G_2}\sum_{G_3} p(G_1)p(X_2|G_2)p(X_3|G_3)p(G_2|G_1)p(G_3|G_1)p(X_1|G_1) \\
&= \sum_{X_1}\sum_{G_2}\sum_{G_3} p(G_1)m_{X_2}(G_2)m_{X_3}(G_3)p(G_2|G_1)p(G_3|G_1)p(X_1|G_1) \\
&= \sum_{G_2}\sum_{G_3} p(G_1)m_{X_2}(G_2)m_{X_3}(G_3)p(G_2|G_1)p(G_3|G_1)
\end{aligned}$$

$$\begin{aligned}
p(G_1 = 1, X_2 = 50, X_3 = 50) = &\sum_{G_2}\sum_{G_3} p(G_1 = 1)p(X_2 = 50|G_2)p(X_3 = 50|G_3)p(G_2|G_1 = 1)p(G_3|G_1 = 1) \\
=&0.5 * [p(X_2 = 50|G_2 = 1)p(X_3 = 50|G_3 = 1)p(G_2 = 1|G_1 = 1)p(G_3 = 1|G_1 = 1) \\
&+ p(X_2 = 50|G_2 = 1)p(X_3 = 50|G_3 = 2)p(G_2 = 1|G_1 = 1)p(G_3 = 2|G_1 = 1) \\
&+ p(X_2 = 50|G_2 = 2)p(X_3 = 50|G_3 = 2)p(G_2 = 2|G_1 = 1)p(G_3 = 2|G_1 = 1) \\
&+ p(X_2 = 50|G_2 = 2)p(X_3 = 50|G_3 = 1)p(G_2 = 2|G_1 = 1)p(G_3 = 1|G_1 = 1)] \\
=&\frac{0.5}{20\pi} * (0.9^2(e^{-\frac{(50-55)^2}{20}})^2 + 0.1^2(e^{-\frac{(50-65)^2}{20}})^2 + 2 * 0.9 * 0.1e^{-\frac{(50-65)^2}{20}}e^{-\frac{(50-55)^2}{20}}) \\
=&0.00529
\end{aligned}$$

$$\begin{aligned}
p(G_1 = 2, X_2 = 50, X_3 = 50) = &\sum_{G_2}\sum_{G_3} p(G_1 = 2)p(X_2 = 50|G_2)p(X_3 = 50|G_3)p(G_2|G_1 = 2)p(G_3|G_1 = 2) \\
=&0.5 * [p(X_2 = 50|G_2 = 1)p(X_3 = 50|G_3 = 1)p(G_2 = 1|G_1 = 2)p(G_3 = 1|G_1 = 2) \\
&+ p(X_2 = 50|G_2 = 1)p(X_3 = 50|G_3 = 2)p(G_2 = 1|G_1 = 2)p(G_3 = 2|G_1 = 2) \\
&+ p(X_2 = 50|G_2 = 2)p(X_3 = 50|G_3 = 2)p(G_2 = 2|G_1 = 2)p(G_3 = 2|G_1 = 2) \\
&+ p(X_2 = 50|G_2 = 2)p(X_3 = 50|G_3 = 1)p(G_2 = 2|G_1 = 2)p(G_3 = 1|G_1 = 2)] \\
=&\frac{0.5}{20\pi} * (0.1^2(e^{-\frac{(50-55)^2}{20}})^2 + 0.9^2(e^{-\frac{(50-65)^2}{20}})^2 + 2 * 0.9 * 0.1e^{-\frac{(50-65)^2}{20}}e^{-\frac{(50-55)^2}{20}}) \\
=&0.00000654
\end{aligned}$$

$$\begin{aligned}
p(G_1|X_2 = 50, X_3 = 50) &= \frac{p(G_1, X_2 = 50, X_3 = 50)}{p(X_2 = 50, X_3 = 50)} \\
p(G_1 = 1|X_2 = 50, X_3 = 50) &= \frac{p(G_1 = 1, X_2 = 50, X_3 = 50)}{p(X_2 = 50, X_3 = 50)} \\
&= \frac{0.00529}{0.00529 + 0.00000654} = 0.988 \\
p(G_1 = 2|X_2 = 50, X_3 = 50) &= \frac{p(G_1 = 2, X_2 = 50, X_3 = 50)}{p(X_2 = 50, X_3 = 50)} \\
&= \frac{0.00000654}{0.00529 + 0.00000654} = 0.012
\end{aligned}$$

Intuitive explanation: If the descendant of G1, that is G1 and G3, are health which means the measure of blood pressure is low, the genotype of G1 is likely a healthy gene.

1.2 Conditional Random Fields

Answer

1. The undirected graph and factor graph of the CRF are shown below:

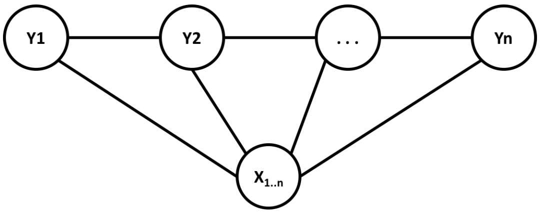

Figure 1.1: undirected graph

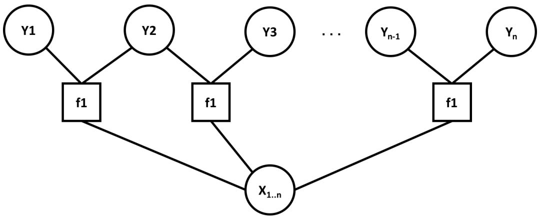

Figure 1.2: factor graph

2. The junction tree of the CRF is shown below:

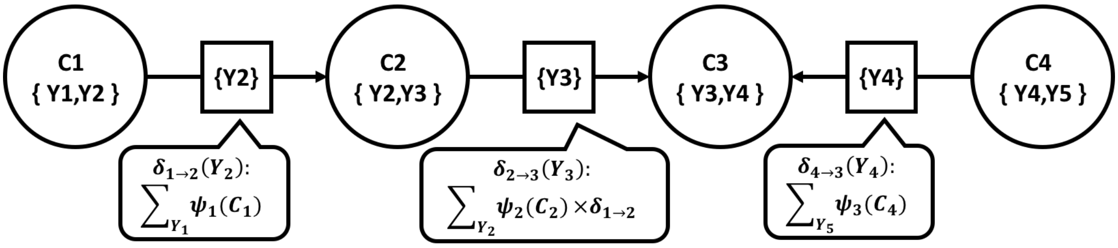

Figure 1.3: junction tree

Answer p(y3|x;w) by running the sum-product algorithm on the junction tree. We choose C3 as root node.

1. In C1: We eliminate Y1 by performing $\sum_{Y_1} \Psi_1(C_1)$.The resulting factor has scope Y2. We send it as a message $\delta_{1\to 2}(Y_2)$ to C2.
2. In C2: We define $\beta_2(Y_2, Y_3) = \delta_{1\to 2}(Y_2) \cdot \Psi_2(Y_2, Y_3)$. We then eliminate Y2 to get a factor over Y3. The resulting factor is $\delta_{2\to 3}(Y_3)$, which is sent to C3.
3. In C4: We eliminate Y5 by performing $\sum_{Y_5} \Psi_3(C_4)$.The resulting factor has scope Y4. We send it as a message $\delta_{4\to 3}(Y_4)$ to C3.
4. In C3: We define $\beta_3(Y_3, Y_4) = \delta_{2\to 3}(Y_3) \cdot \delta_{4\to 3}(Y_4) \cdot \Psi_3(Y_3, Y_4)$

The factor β_3 is a factor over Y3, Y4 that encodes the joint distribution $P(y_3, y_4|x;w)$: all the CPDs have been multiplied in, and all the other variables have been eliminated. If we now want to obtain $P(y_3|x;w)$:, we simply sum out y_4.
Now we know that:

$$\Psi_1(C_1) = \exp\{w^T f(x, y_1, y_2)\} \quad \Psi_2(C_2) = \exp\{w^T f(x, y_2, y_3)\}$$
$$\Psi_3(C_3) = \exp\{w^T f(x, y_3, y_4)\} \quad \Psi_4(C_4) = \exp\{w^T f(x, y_4, y_5)\}$$

$$\delta_{1\to 2}(Y_2) = \sum_{y_1} exp\{w^T f(x, y_1, y_2)\}$$
$$\delta_{2\to 3}(Y_3) = \sum_{y_2} exp\{w^T f(x, y_2, y_3)\} \cdot \delta_{1\to 2}(Y_2)$$
$$\delta_{4\to 3}(Y_4) = \sum_{y_5} exp\{w^T f(x, y_4, y_5)\}$$

$$\begin{aligned}
\beta_3(Y_3, Y_4) &= \delta_{2\to 3}(Y_3) \cdot \delta_{4\to 3}(Y_4) \cdot \Psi_3(Y_3, Y_4) \\
&= \sum_{y_2} exp\{w^T f(x, y_2, y_3)\} \cdot \sum_{y_1} exp\{w^T f(x, y_1, y_2)\} \cdot \sum_{y_5} exp\{w^T f(x, y_4, y_5)\} \cdot \Psi_3(Y_3, Y_4) \\
p(y_3|x; w) &= \frac{1}{Z(x; w)} \sum_{y_4} \beta_3(Y_3, Y_4)
\end{aligned}$$

3. **Proof**:

$$\begin{aligned}
\frac{\partial L}{\partial w} &= \frac{\partial \sum_{(x,y)\in D} \{w^T \sum_{i=2}^n f(x, y_i, y_{i-1}) - logZ(x; w)\}}{\partial w} \\
&= \sum_{(x,y)\in D} \sum_{i=2}^n f(x, y_i, y_{i-1}) - \sum_{(x,y)\in D} \frac{\partial logZ(x; w)}{\partial w}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial logZ(x; w)}{\partial w} &= \frac{1}{Z(x; w)} \frac{\partial Z(x; w)}{\partial w} \\
&= \frac{1}{Z(x; w)} \frac{\partial \sum_{y'\in L^n} exp\{w^T \sum_{i=2}^n f(x, y'_i, y'_{i-1})\}}{\partial w} \\
&= \sum_{y'\in L^n} \frac{1}{Z(x; w)} exp\{w^T \sum_{i=2}^n f(x, y'_i, y'_{i-1})\} \sum_{i=2}^n f(x, y'_i, y'_{i-1}) \\
&= \sum_{y'\in L^n} \sum_{i=2}^n p(y|x; w) f(x, y'_i, y'_{i-1}) \\
&= \sum_{i=2}^n E_{p(y'|x;w)} f(x, y'_i, y'_{i-1})
\end{aligned}$$

Thus

$$\frac{\partial L}{\partial w} = \sum_{(x,y)\in D} \sum_{i=2}^n \{f(x, y_i, y_{i-1}) - E_{p(y|x;w)}[f(x, y'_i, y'_{i-1})]\}$$

4.

$$\sum_{i=2}^{n} \mathbb{E}_{p(\mathbf{y'}|x;w)}[f(x,y'_i,y'_{i-1})] = \sum_{i=2}^{n}\sum_{y'} p(\mathbf{y'}|x;w) f(x,y'_i,y'_{i-1})$$
$$= \sum_{i=2}^{n}\sum_{y'_i,y'_{i-1}} p(y'_i,y'_{i-1}|x;w) f(x,y'_i,y'_{i-1})$$

Belief propagation algorithm on the junction tree:
Because it's easy to get $f(x,y'_i,y'_{i-1})$, we only need to compute$p(y'_i,y'_{i-1}|x;w)$.
Step 1: For each node, multiply all the messages and divide by the one that is coming from node we are sending the message to. The same as VE:

$$\delta_{i\to j} = \sum_{C_i - S_{ij}} \Pi_{k\in N(i)/j}\delta_{k\to i}$$

Initialize the messages on the edges to 1 and potentials $\pi_i^0 = \exp(w^T f(x,y_i,y_{i-1}))$, i=2..n
Step 2: Store the last message on the edge and divide each passing message by the last stored. choose the last clique as root node, pass message from C_2 to C_n.

$$\delta_{i-1\to i} = \sum_{y_{i-2}} \pi_{i-1}^1(y_{i-2},y_{i-1})$$

where $\pi_{i-1}^1(y_{i-2},y_{i-1}) = \pi_{i-1}^0(y_{i-2},y_{i-1})\delta_{i-1\to i}$
Step 3: pass message from C_n to C_2.

$$\delta_{i\to i-1} = \frac{\sum_{y_i}\pi_i^1(y_{i-1},y_i)}{\delta_{i-1\to i}}$$

where $\mu_{i-1,i} = \pi_{i-1}^1(y_{i-2},y_{i-1}) = \delta_{i-1\to i}\pi_{i-1}^1(y_{i-2},y_{i-1})$
Step 4: Finally, we get each clique's $\pi_i(y_{i-1},y_i)$. Thus

$$p(y'_i,y'_{i-1}|x;w) = \frac{\pi_i^1(y'_{i-1},y'_i)}{\sum_{y'_{i-1},y'_i}\pi_i^1(y'_{i-1},y'_i)}$$

2 Deep Generative Models: Class-conditioned VAE

1.

$$\begin{aligned}
\log p(\mathbf{x}|y,\theta) &= \log \int p(\mathbf{x},\mathbf{z}|y,\theta)\mathbf{dz} \\
&= \log \int \frac{p(\mathbf{x},\mathbf{z}|y,\theta)}{q(\mathbf{z}|\mathbf{x},y,\phi)} q(\mathbf{z}|\mathbf{x},y,\phi)\mathbf{dz} \\
&\geq \int q(\mathbf{z}|\mathbf{x},y,\phi) \log \frac{p(\mathbf{x},\mathbf{z}|y,\theta)}{q(\mathbf{z}|\mathbf{x},y,\phi)}\mathbf{dz} \\
&= \int q(\mathbf{z}|\mathbf{x},y,\phi) \log \frac{p(\mathbf{x}|\mathbf{z},y,\theta)\,p(\mathbf{z}|y,\theta)}{q(\mathbf{z}|\mathbf{x},y,\phi)}\mathbf{dz} \\
&= \int q(\mathbf{z}|\mathbf{x},y,\phi) \log p(\mathbf{x}|\mathbf{z},y,\theta)\mathbf{dz} - \int q(\mathbf{z}|\mathbf{x},y,\phi)[\log q(\mathbf{z}|\mathbf{x},y,\phi) - \log p(\mathbf{z}|y,\theta)]\mathbf{dz} \\
&= \mathbb{E}_{q(\mathbf{z}|\mathbf{x},y,\phi)}[\log p(\mathbf{x}|\mathbf{z},y,\theta)] - KL(q(\mathbf{z}|\mathbf{x},y,\phi)\,\|\,p(\mathbf{z}|y,\theta)) \\
&= L(\theta,\phi,\mathbf{x},y)
\end{aligned}$$

2. The implementation of the algorithm with ZhuSuan is in the file hw4_VAE.ipynb/hw_VAE.html
3. Visualize generations of model.

Figure 2.1: VAE

Acknowledgement
Thanks greatly for my roommate Shuya Li's help.