Solution of Assignment 4

January 12, 2016

1. In a second-order Markov chain, each state depends on the two previous states, i.e., $P\left[X^{(t+1)}=S_k|X^{(t)}=S_j,X^{(t-1)}=S_i,\cdots\right]=P\left[X^{(t+1)}=S_k|X^{(t)}=S_j,X^{(t-1)}=S_i\right]$. Show that the second-order Markov chain can always be converted to a first-order Markov chain. (Hint: by redesign the states)

   Answer:
   Denote $p(w|y,z)=P\left[X^{(t+1)}=w|X^{(t)}=y,X^{(t-1)}=z\right]$
   Let $Z^{(t)}=(X^{(t)},X^{(t+1)})$
   We have $P\left[Z^{(t+1)}=(w,w')|Z^{(t)}=(y,z)\right]=\begin{cases}0 & if\, w'\neq y\\ p(w|y,z) & otherwise\end{cases}$

2. Prove that a Bayesian network must be a Directed Acyclic Graph (DAG).

   Answer:
   We proof by induction by considering Bayesian networks with i nodes, $i\geq 2$. When $i=2$, we have either $P(X_1,X_2)=P(X_1|X_2)P(X_2)$ or $P(X_1,X_2)=P(X_2|X_1)P(X_1)$, which all results in a DAG. Now suppose that every Bayesian network with i nodes is a DAG, and consider a Bayesian network with $i+1$ nodes given by $P(X_1,X_2,\cdots,X_{i+1})=\prod_{j=1}^{i+1}P(X_j|parent(X_j))$. Let L be a node in the network that has no child. Then the subgraph corresponding to $\{X_1,X_2,\cdots,X_{i+1}\}\backslash L$ is a DAG based on our assumption. Since all links between nodes in the subgraph and L have the same direction pointing to L, the Bayesian network cannot have any path passing through L and therefore must still be a DAG.

3. Given random variables A, B, C, and D, answer true or false and justify your answer:

   (a) $\{A\}\perp\!\!\!\perp\{B\}|\{C\}$ implies $\{A\}\perp\!\!\!\perp\{B\}$;
   (b) $\{A\}\perp\!\!\!\perp\{B\}$ implies $\{A\}\perp\!\!\!\perp\{B\}|\{C\}$;
   (c) $\{A\}\perp\!\!\!\perp\{B,C\}|\{D\}$ implies $\{A\}\perp\!\!\!\perp\{B\}|\{D\}$.

   Answer:

   (a) False, as $p(A,B)=\int p(A,B,C)dC=\int p(A,B|C)p(C)dC=\int p(A|C)p(B|C)p(C)dC$ does not generally equal to $p(A)p(B)$ for all distributions.
   (b) False, as $p(A,B|C)=\frac{p(C|A,B)p(A,B)}{p(C)}=\frac{p(C|A,B)p(A)p(B)}{p(C)}$ does not generally equal to $p(A|C)p(B|C)=\frac{p(A|C)p(C|B)p(B)}{p(C)}$.
   (c) True, since $p(A,B|D)=\int p(A,B,C|D)dC=\int p(A|D)p(B,C|D)dC=p(A|D)\int p(B,C|D)dC=p(A|D)p(B|D)$.

4. Given a Hidden Markov model with time homogeneous Gaussian emission probability $P[\boldsymbol{x}^{(t)}|z_i^{(t)}, \theta_i] = \frac{1}{(2\pi)^{d/2}det(\boldsymbol{\Sigma}_i)^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}^{(t)}-\boldsymbol{\mu}_i)^\top \boldsymbol{\Sigma}_i^{-1}(\boldsymbol{x}^{(t)}-\boldsymbol{\mu}_i)}$, where $\theta_i = (\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$. Consider the problem finding $\Theta = (\boldsymbol{\pi}^{(1)}, \boldsymbol{A}, \{\theta_k\}_{k=1}^K)$ using the EM algorithm. Show that maximizing $\mathcal{Q}(\Theta; \Theta^{old})$ in the M-step gives $\boldsymbol{\mu}_i = \frac{\sum_{t=1}^T \gamma_i^{(t)} \boldsymbol{x}^{(t)}}{\sum_{t=1}^T \gamma_i^{(t)}}$ and $\boldsymbol{\Sigma}_i = \frac{\sum_{t=1}^T \gamma_i^{(t)}(\boldsymbol{x}^{(t)}-\boldsymbol{\mu}_i)(\boldsymbol{x}^{(t)}-\boldsymbol{\mu}_i)^\top}{\sum_{t=1}^T \gamma_i^{(t)}}$.

Answer:
$\mathcal{Q}(\Theta; \Theta^{old}) = \sum_{i=1}^k \ln(\pi_i^{(1)}) r_i^{(1)} + \sum_{t=1}^{T-1} \sum_{i=1}^K \sum_{j=1}^K \ln(a_{i,j}) \xi_{i,j}^{(t)} + \sum_{t=1}^T \sum_{i=1}^K \ln(P[x^{(t)}|z_i^{(t)}, \theta_i]) \gamma_i^{(t)}$
s.t. $\sum_{t=1}^T \pi_i^{(1)} = 1$, $\sum_{j=1}^K a_{i,j} = 1$ for all $1 \le i \le K$
solve $\pi^{(i)}$: $L(\pi^{(1)}, \alpha) = \sum_{i=1}^k \ln(\pi_i^{(1)}) \gamma_i^{(1)} - \alpha(\sum_{i=1}^k \pi_i^{(1)} - 1)$
solve A : $L(A, \{\alpha_i\}_{i=1}^k) = \sum_{i=1}^{T-1} \sum_{i=1}^K \sum_{j=1}^K \ln(a_{i,j}) \xi_{i,j}^{(t)} - \sum_{i=1}^K \alpha_i (\sum_{j=1}^k a_{i,j} - 1)$
sovle $\{\theta_i\}_{i=1}^k$: max $\sum_{t=1}^T \sum_{i=1}^K \ln(P[x^{(t)}|z_i^{(t)}, \theta_i]) \gamma_i^{(t)}$, assume $P[x^{(t)}|z_i^{(t)}, \theta_i] \sim N(\vec{\mu_i}, \Sigma_i)$
$f = \sum_{t=1}^T \sum_{i=1}^K \ln(P[x^{(t)}|z_i^{(t)}, \theta_i]) \gamma_i^{(t)} =$
$-\frac{1}{2} \sum_{t=1}^T \sum_{i=1}^K (x^{(t)} - \mu_i)^\top \Sigma_i^{-1} (x^{(t)} - \mu_i) \gamma_i^{(t)} - \sum_{t=1}^T \sum_{i=1}^K \ln(2\pi^{\frac{d}{2}} \det(\Sigma_i)^{\frac{1}{2}}) \gamma_i^{(t)}$
$\frac{\partial f}{\partial \mu_i} = -\frac{1}{2} (\sum_{t=1}^T \gamma_i^{(t)} (x^{(t)} - \mu_i)^\top \cdot 2\Sigma_i^{-1}) = 0 \Rightarrow \mu_i = \frac{\sum_{t=1}^T \gamma_i^{(t)} x^{(t)}}{\sum_{t=1}^T \gamma_i^{(t)}}$
$\frac{\partial f}{\partial \Sigma_i^{-1}} = -\frac{1}{2} \sum_{t=1}^T (x^{(t)} - \mu_i)^\top (x^{(t)} - \mu_i) \gamma_i^{(t)} + \frac{1}{2} \sum_{t=1}^T \gamma_i^{(t)} \Sigma_i = 0 \Rightarrow \Sigma_i = \frac{\sum_{t=1}^T (x^{(t)} - \mu_i)^\top (x^{(t)} - \mu_i) \gamma_i^{(t)}}{\sum_{t=1}^T \gamma_i^{(t)}}$