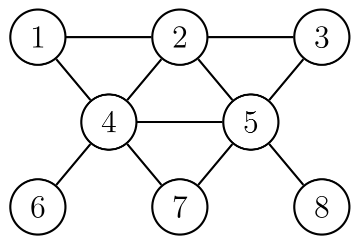
STA531 Midterm Exam 2 Solutions

1. Graphical models

   (a) $p(x_1, \ldots, x_8) = p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2)p(x_5|x_2, x_3)p(x_6|x_4)p(x_7|x_4, x_5)p(x_8|x_5)$

   (b) Yes, Indeterminate, Yes, Indeterminate

   (c)



2. Markov chains and graphical models

   (a) Suppose p respect G_{D}. Then

$$p(x_1, \ldots, x_n) = p(x_1) \prod_{i=1}^{n-1} p(x_{i+1}|x_i) = \prod_{i=1}^{n-1} \varphi_i(x_i, x_{i+1})$$

   if we define $\varphi_1(x_1, x_2) = p(x_1)p(x_2|x_1)$ and $\varphi_i(x_i, x_{i+1}) = p(x_{i+1}|x_i)$ for $i = 2, \ldots, n-1$. The functions $\varphi_1, \varphi_2, \ldots, \varphi_{n-1}$ are nonnegative, and the maximal cliques of G_{U} are $\{1, 2\}, \{2, 3\}, \ldots, \{n-1, n\}$. Therefore, p respects G_{U}.

   (b) Suppose p respects G_{U}. Then

$$p(x_1, \ldots, x_n) = p(x_1) \prod_{i=2}^{n} p(x_i|x_1, \ldots, x_{i-1}) = p(x_1) \prod_{i=2}^{n} p(x_i|x_{i-1}) = \prod_{i=1}^{n} p(x_i|x_{\text{pa}(i)})$$

   since

   - $X_i \perp X_1, \ldots, X_{i-2} \mid X_{i-1}$ by the separation criterion for undirected graphical models, because all paths from i to $\{1, \ldots, i-2\}$ pass through $i-1$, and
   - $\text{pa}(1) = \varnothing$ and $\text{pa}(i) = \{i-1\}$ for $i = 2, \ldots, n$.

   Therefore, p respects G_{D}.

3. Modeling the data collection process

   As usual, the appropriate posterior on θ to use is $p(\theta|y_{\text{obs}}, I)$, since this is the posterior obtained by conditioning on all of the observations. Since $I_j = 1$ for all $j = 1, \ldots, n$, obs $= \{1, \ldots, n\}$. Further, the distribution of $I|y, \theta$ does not depend on θ (because $p(I|y, \theta) = \prod_{j=1}^n \phi_{y_j}^{I_j}(1-\phi_{y_j})^{1-I_j} \propto_\theta 1$). Consequently, the data collection process plays no role:

$$\begin{aligned} p(\theta|y_{\text{obs}}, I) = p(\theta|y, I) \propto p(y, I, \theta) &= p(\theta)p(y|\theta)p(I|y, \theta) \\ &\propto_\theta p(\theta)p(y|\theta) \propto_\theta \theta^{a-1}(1-\theta)^{b-1} \prod_{j=1}^{n} \theta^{y_j}(1-\theta)^{1-y_j} \\ &\propto_\theta \text{Beta}(\theta \mid a + \textstyle\sum y_j,\ b + n - \sum y_j). \end{aligned}$$

4. Viterbi algorithm

Define $\nu_n(z_n) = 1$ for each $z_n = 1, \ldots, m$, and denote $p(z_1) = p(z_1|z_0)$ for notational simplicity. Then

$$
\begin{aligned}
\max_{z_{1:n}} p(x_{1:n}, z_{1:n}) &= \max_{z_{1:n}} \prod_{t=1}^{n} p(z_t|z_{t-1})p(x_t|z_t) \\
&\vdots \\
&= \max_{z_{1:j}} \left(\prod_{t=1}^{j} p(z_t|z_{t-1})p(x_t|z_t) \right) \underbrace{\max_{z_{j+1}} p(z_{j+1}|z_j)p(x_{j+1}|z_{j+1})\nu_{j+1}(z_{j+1})}_{\text{call this } \nu_j(z_j)} \\
&\vdots \\
&= \max_{z_1} p(z_1)p(x_1|z_1)\nu_1(z_1).
\end{aligned}
$$

Therefore, we have the following algorithm:

(1) Set $\nu_n(z_n) = 1$ for $z_n = 1, \ldots, m$.

(2) For each $j = n-1, n-2, \ldots, 1$, for each $z_j = 1, \ldots, m$, compute

$$\nu_j(z_j) = \max_{z_{j+1}} p(z_{j+1}|z_j)p(x_{j+1}|z_{j+1})\nu_{j+1}(z_{j+1}).$$

(3) Return $\max_{z_1} p(z_1)p(x_1|z_1)\nu_1(z_1)$.

(Note: If you showed that your algorithm takes nm^2 time, then that is great, but this was not required—as long as your algorithm actually does take nm^2 time, then that is sufficient.)