Machine Learning 2 - Homework Week 2

Minh Ngo
MSc Artificial Intelligence
University of Amsterdam
nlminhtl@gmail.com

April 12, 2016

Collaborators: Arthur Bražinskas, Riaan Zoetmulder

Problem 3

3.1

$$p(x) = \mathcal{N}(x|\mu, \Sigma) \Leftrightarrow x \sim \mathcal{N}(x|\mu, \Sigma) \tag{1}$$

$$q(x) = \mathcal{N}(x|m, L) \tag{2}$$

$$
\begin{aligned}
\mathcal{KL}(p \parallel q) &= -\int p(x) \ln \frac{q(x)}{p(x)} dx \\
&= -\int p(x) \Big[\ln q(x) - \ln p(x) \Big] dx \\
&= -\int p(x) \ln q(x) dx + \int p(x) \ln p(x) dx \\
&= -\int p(x) \ln \Bigg[\frac{1}{(2\pi)^{D/2}|L|^{1/2}} \exp\Bigg(-\frac{1}{2}(x-m)^T L^{-1}(x-m) \Bigg) \Bigg] dx \\
&\quad + \int p(x) \ln \Bigg[\frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\Bigg(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu) \Bigg) \Bigg] dx \\
&= \frac{1}{2} \ln \frac{|L|}{|\Sigma|} \underbrace{\int p(x) dx}_{=1} + \frac{1}{2} \int (x-m)^T L^{-1}(x-m) p(x) dx \\
&\quad - \frac{1}{2} \int (x-\mu)^T \Sigma^{-1}(x-\mu) p(x) dx
\end{aligned}
\tag{3}
$$

From the law of unconscious statistician we have:

$$
\mathcal{KL}(p \parallel q) = \frac{1}{2} \ln \frac{|L|}{|\Sigma|} + \frac{1}{2} E[(x-m)^T L^{-1}(x-m)] - \frac{1}{2} E[(x-\mu)^T \Sigma^{-1}(x-\mu)] \tag{4}
$$

Applying Matrix Cookbook (380):

$$
\begin{aligned}
\mathcal{KL}(p \parallel q) &= \frac{1}{2} \ln \frac{|L|}{|\Sigma|} + \frac{1}{2} \Big[(\mu-m)^T L^{-1}(\mu-m) + Tr(L^{-1}\Sigma) \Big] \\
&\quad - \frac{1}{2} \Big[(\mu-\mu)^T \Sigma^{-1}(\mu-\mu) + Tr(\Sigma^{-1}\Sigma) \Big] \\
&= \frac{1}{2} \ln \frac{|L|}{|\Sigma|} + \frac{1}{2}(\mu-m)^T L^{-1}(\mu-m) + \frac{1}{2} Tr(L^{-1}\Sigma) - \frac{1}{2} D
\end{aligned}
\tag{5}
$$

3.2

$$
\begin{aligned}
\mathcal{H}(x) &= -\int p(x)\ln p(x)dx \\
&= -\int \ln \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} \exp\Big(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\Big)p(x)dx \\
&= -\int \Big[\frac{D}{2}\ln(2\pi) + \frac{1}{2}\ln|\Sigma| + \frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\Big]p(x)dx \\
&= \frac{D}{2}\ln(2\pi) + \frac{1}{2}\ln|\Sigma| + \frac{1}{2}E[(x-\mu)^T\Sigma^{-1}(x-\mu)]
\end{aligned}
\tag{6}
$$

Applying Matrix Cookbook (380):

$$
\begin{aligned}
\mathcal{H}(x) &= \frac{D}{2}\ln(2\pi) + \frac{1}{2}\ln|\Sigma| + \frac{1}{2}(\mu-\mu)^T\Sigma^{-1}(\mu-\mu) + \frac{1}{2}Tr(\Sigma^{-1}\Sigma) \\
&= \frac{D}{2}\ln(2\pi) + \frac{1}{2}\ln|\Sigma| + \frac{1}{2}D
\end{aligned}
\tag{7}
$$

Problem 4

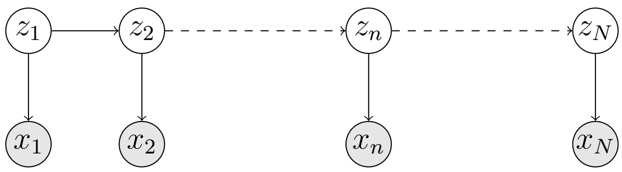

Figure 1: Markov chain of latent variables.

1. The arrows on the path from $z_{n-1} \to z_n \to x_n$ meet head-to-tail $\Rightarrow$ $x_1, .., x_{n-1}$ are d-separated from x_n given $z_n \Rightarrow x_n \perp\!\!\!\perp x_1, .., x_{n-1}|z_n$.

   Consequently: $p(x_1, .., x_{n-1}|x_n, z_n) = p(x_1, .., x_{n-1}|z_n)$ q.e.d.

2. Tail-to-tail path $x_{n-k} \leftarrow z_{n-k} \to z_n$ where $k \geq 1 \Rightarrow z_n \perp\!\!\!\perp x_1, .., x_{n-1}|z_{n-1}$. Consequently: $p(x_1, .., x_{n-1}|z_{n-1}, z_n) = p(x_1, .., x_{n-1}|z_{n-1})$ q.e.d.

3. Head-to-tail path $z_n \to z_{n+1} - - > x_{n+k}$ where $k \geq 1$. Consequently, $z_n \perp\!\!\!\perp x_{n+1}, .., x_N|z_{n+1}$ and $p(x_{n+1}, .., x_N|z_n, z_{n+1}) = p(x_{n+1}, .., x_N|z_{n+1})$ q.e.d.

4. Tail-to-tail path $x_N \leftarrow z_N \to z_{N+1}$ that blocks all previous x_k where $k \geq 0$. Consequently $z_{N+1} \perp\!\!\!\perp \mathbf{X}|z_N$ and $p(z_{N+1}|z_N, \mathbf{X}) = p(z_{N+1}|z_N)$ where $\mathbf{X} = \{x_1, .., x_N\}$ q.e.d.

Problem 5

5.1

Let $z_1, .., z_N$ be common parents of nodes X and Y and the edge $\{X, Y\}$ be a covered edge. To prove that after reversing the edge $\{X, Y\}$ all independent relations stay the same we have to prove that all conditional independence will be the same given X or Y.

1. Let K_i and K_j be some descendant of X (they may be not direct children).
   - $z_k \to X \to K_i$ forms a head-to-tail path and will stay the same after reversing (is blocked given X).
   - $K_i \leftarrow X \to Y$ forms tail-to-tail path and will become head-to-tail relation (that are blocked given X in both cases).
   - $K_i \leftarrow X \to K_j$ forms tail-to-tail path and will remain unchanged after reversing $\{X, Y\}$.
2. We are considering descendants of Y in the same way:
   - $z_k \to Y \to K_i$ remains blocked given X.
   - $X \to Y \to K_i$ is head-to-tail path and will become tail-to-tail after reversing and remains blocked given X.
   - $K_i \leftarrow Y \to K_j$ is tail-to-tail path and will remain unchanged after reversing $\{X, Y\}$.

Consequently $\mathcal{G}$ and $\mathcal{G}'$ encode the same set of independent relations q.e.d.

5.2

Counterexample can be a graph $z \to y \leftarrow x$, where $z \not\perp\!\!\!\perp x|y$ (Bishop 8.29). After reversing the edge we get a graph $z \to y \to x$ that makes $z \perp\!\!\!\perp x|y$ (according to the d-separation criteria).