Statistical learning with and from graphs

Marco Corneli

---

1. Consider Figure 2.9 at page 20 in Ch. 2 of the lecture notes. Write down the factorization of a joint probability p_θ that is Markovian with respect to the DAG in that figure.
2. You need to have understood the appendix at the end of Chapter 2 in order to solve this. Consider three random variables (X, Y, Z) having the following distribution:

$$X \sim \mathcal{N}(0, 1)$$
$$Y|X = x \sim \mathcal{N}(\alpha x, 1)$$
$$Z|X = x, Y = y \sim \mathcal{N}(\beta y + \gamma x, 1)$$

   (a) Sketch the DAG associated with the joint density $p_\theta(x, y, z)$ (where $\theta := \{\alpha, \beta, \gamma\}$.)

   (b) Find an explicit expression for $p_\theta(z|y) = \int_{\mathbb{R}} p_\theta(z, x|y)dx$ (**Hint:** Everything is Gaussian here, so in order to specify a pdf you need to compute a mean and a variance.) and compute $\mathbb{E}(Z|Y = y)$.

   (c) Find an explicit expression for $p_\theta(z|Y := y)$ and compute $\mathbb{E}(Z|Y := y) = \int_{\mathbb{R}} zp_\theta(z|Y := y)dz$. Compare with the previous point.

   (d) Find the joint distribution of (Y, Z). What is the correlation ρ between Y and Z ?

   (e) Suppose that X is not observed, in other term suppose it is a latent *confounding* variable. Suppose we declare that Y causes Z if $\rho \neq 0$ and Y does not cause Z if $\rho = 0$. Show that this will lead to erroneous conclusions.

   (f) Suppose finally that a randomized experiment is conducted breaking the tie from X to Y, in more details assume that

$$X \sim \mathcal{N}(0,1)$$
$$Y|X = x \sim \mathcal{N}(\alpha, 1)$$
$$Z|X = x, Y = y \sim \mathcal{N}(\beta y + \gamma x, 1)$$

Show that now the same declaration "Y causes Z if $\rho \neq 0$ and Y does not cause Z if $\rho = 0$" is now correct.

3. Prove Proposition 2 at page 33 (Ch.3) of the lecture notes.
4. Still in the context of Gaussian Mixture models (Ch.3) consider the lower bound in Eq.(3.8). Now instead of fixing $q(\cdot) = p_{\mathbf{z}}$ as we did in class, assume that

$$q(Z) = \prod_{i=1}^{N} q(Z_i) = \prod_{i=1}^{N} \tau_{iZ_{ik}} = \prod_{i=1}^{N}\prod_{k=1}^{K} \tau_{ik}^{Z_{ik}}$$

where $\tau_{i1}, \ldots, \tau_{iK}$ are unknown, positive and $\sum_{k=1}^{K} = 1$, for all i. This is another way to state that we assume that under a probability associated with the probability mass function $q(\cdot)$, let us call it $\mathbb{Q}$, then $\mathbb{Q}(Z_{ik} = 1) = \tau_{ik}$ independently for all i.

   (a) Compute the lower bound $\mathcal{L}(q, \Theta) := \mathbb{E}_{Z \sim q}\left[\log\left(\frac{p(\mathbf{x}, \mathbf{Z}|\Theta)}{q(\mathbf{Z})}\right)\right]$, where $q(Z)$ is defined above and $\log p(\mathbf{x}, \mathbf{Z}|\Theta)$ is the one in Eq.(3.6), page 30 of your notes.

   (b) For a fixed pair (i', k'), maximize the $\mathcal{L}(q, \Theta)$ that you obtained with respect to $\tau_{i'k'}$ under the constraint $\sum_{k=1}^{K} \tau_{i'k} = 1$ (**Hint:** you might wish to adopt a Lagrange multiplier).

   (c) Compare the optimal $\tau_{i'k'}$ that you found with the one in Eq. (3.10). What can you see? Conclude.