Statistical learning with and from graphs

Marco Corneli

3h

---

1. We saw in class the following penalized log-likelihood

$$l_R(\beta) = \sum_{i=1}^{N} \left(y_i X_i \beta - \log(1 + e^{X_i \beta}) \right) - \frac{\lambda}{2} ||\beta||_2$$

   corresponding to a penalized logistic regression where $y_1, \ldots, y_N$ are the observed binary labels, $X_1, \ldots, X_N \in \mathbb{R}^D$ are the corresponding feature vectors, $\beta \in \mathbb{R}^D$ is an unknown parameter to estimate and λ is a positive real constant, assumed to be known.

   (a) Compute the gradient $\nabla_\beta l_R(\beta)$.

   (b) Recalling that the Heissian matrix corresponding to the above loss is

$$H_\beta l_R(\beta) = -\sum_{i=1}^{N} \left(p_i(1-p_i) X_i^T X_i \right) - \lambda I_D$$

   where $p_i = 1/(1 + e^{-X_i \beta})$ show that it is negative definite. What does it mean?

   (c) Write down the Newton-Raphson update for β.

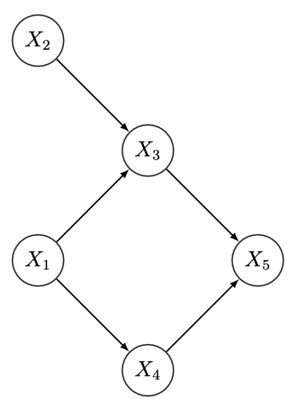

Figure 2.9: A slightly more difficult DAG.

2. Look at Figure 2.9.

   (a) Write down the factorization of a joint probability p_θ that is Markovian with respect to the DAG in that figure.

   (b) Consider the sub-graph only containing (X_1, X_2, X_3), with X_3 being a collider. Show that X_1 and X_2 are independent but they are conditionally dependent given X_3.

3. In the context of mixture models, we are given the log-likelihood of the observed data $l(\Theta) := \log p(\mathbf{x}|\Theta) = \log\left(\sum_{\mathbf{z}} p(\mathbf{x}, \mathbf{z}|\Theta)\right)$, where $\mathbf{x} := (x_1, \ldots, x_N)$ are the observations, $\mathbf{z} := (z_1, \ldots, z_N)$ are the latent variables labeling the clusters' membership and Θ are the model parameters. Show that

$$l(\Theta) = \mathcal{L}(q, \Theta) + \mathbb{E}_q \left[\log\left(\frac{q(\mathbf{Z})}{p(\mathbf{Z}|\mathbf{x}, \Theta)}\right)\right]$$

where $\mathcal{L}(q, \Theta) := \mathbb{E}_q \left[\log\left(\frac{p(\mathbf{x}, \mathbf{Z}|\Theta)}{q(\mathbf{Z})}\right)\right]$ and $q(\cdot)$ is any distribution on $\mathbf{Z}$ with a proper support.

4. We saw in class the stochastic block model for undirected binary graphs, whose log-likelihood of the complete data is

$$\log p(A, \mathbf{z}|\Pi, \alpha) = \frac{1}{2} \sum_{j \neq i}^{N} \sum_{k,l}^{K} \left(z_{ik} z_{jl} A_{jl} \log \pi_{kl} + z_{ik} z_{jl} (1 - A_{ij}) \log(1 - \pi_{kl})\right) + \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \log \alpha_k$$

   (a) Recall the mains assumptions in this model.
   (b) For a fixed pair (k, l), compute the Maximum Likelihood estimates of π_{kl} and α_k from the log-likelihood above.
   (c) Show that $A_{ij}|\{Z_i = k\}$ follows a Bernoulli distribution with parameter $\overline{\pi}_k := \sum_{l=1}^{K} \pi_{kl} \alpha_l$.
   (d) Show that, given $\{Z_i = k\}$, then A_{ij_1} is independent from A_{ij_2}, for all $j_1 \neq j_2$[1].
   (e) According to previous points (c) and (d), what can you conclude about the distribution of the *degree* $d_i := \sum_{j=1}^{N} A_{ij}$?

[1] *Hint:* You might focus on $\mathbb{E}[A_{ij_1} A_{ij_2}|Z_i = k]$ and compare it with...