Tutorial Week 9

COMP9418 – Advanced Topics in Statistical Machine Learning, 17s2, UNSW Sydney

Instructor: Edwin V. Bonilla

Last Update: Wednesday 13th September, 2017 at 10:45

1. Consider an exponential family distribution:

$$p(\mathbf{y}|\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})} \exp\left(\boldsymbol{\theta}^T \boldsymbol{\phi}(\mathbf{y})\right) \tag{1}$$

$$= \exp\left(\boldsymbol{\theta}^T \boldsymbol{\phi}(\mathbf{y}) - A(\boldsymbol{\theta})\right), \tag{2}$$

where $\boldsymbol{\theta}$ are the parameters of the distribution and $A(\boldsymbol{\theta}) = \log Z(\boldsymbol{\theta})$ is the log partition function, with $Z(\boldsymbol{\theta}) = \sum_{\mathbf{y}} \exp\left(\boldsymbol{\theta}^T \boldsymbol{\phi}(\mathbf{y})\right)$ for discrete $\mathbf{y}$.

   (a) Show that

$$\frac{dA}{d\boldsymbol{\theta}} = \mathbb{E}_{p(\mathbf{y}|\boldsymbol{\theta})}[\boldsymbol{\phi}(\mathbf{y})] = \sum_{\mathbf{y}} p(\mathbf{y}|\boldsymbol{\theta})\boldsymbol{\phi}(\mathbf{y}). \tag{3}$$

   (b) Show that

$$\frac{dA}{d\boldsymbol{\theta} d\boldsymbol{\theta}^T} = \mathbb{C}\mathrm{ov}[\boldsymbol{\phi}(\mathbf{y})] = \mathbb{E}_{p(\mathbf{y}|\boldsymbol{\theta})}\left[\boldsymbol{\phi}(\mathbf{y})\boldsymbol{\phi}(\mathbf{y})^T\right] - \mathbb{E}_{p(\mathbf{y}|\boldsymbol{\theta})}[\boldsymbol{\phi}(\mathbf{y})]\mathbb{E}_{p(\mathbf{y}|\boldsymbol{\theta})}[\boldsymbol{\phi}(\mathbf{y})]^T. \tag{4}$$

   (c) Given training data $\mathcal{D} = \{\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(N)}\}$ show that the optimum of the average log likelihood:

$$\mathcal{L}(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{n=1}^{N} \log p(\mathbf{y}^{(n)}|\boldsymbol{\theta}) \tag{5}$$

   is achieved when:

$$\mathbb{E}_{p_{\text{emp}}}[\boldsymbol{\phi}(\mathbf{y})] = \mathbb{E}_{p(\mathbf{y}|\boldsymbol{\theta})}[\boldsymbol{\phi}(\mathbf{y})], \tag{6}$$

   where $\mathbb{E}_{p_{\text{emp}}}[\boldsymbol{\phi}(\mathbf{y})]$ denotes the empirical expectation of $\boldsymbol{\phi}(\mathbf{y})$.

   (d) Show that a tabular MRF is an exponential family distribution, specifying what the parameter vector $\boldsymbol{\theta}$ and feature vector $\boldsymbol{\phi}$ are. What is $\frac{dA}{d\boldsymbol{\theta}}$ and what does Equation (6) imply in this case? How does this relate to the MLE for directed graphical models?

2. Consider the graphical model in Figure 1.

   (a) Give the Markov blanket for every node in the graph.

   (b) Confirm or refute the following conditional independence statements:

      i. $x_1 \perp\!\!\!\perp x_3 | x_2$

      ii. $x_1 \perp\!\!\!\perp x_3 | x_2, x_4$

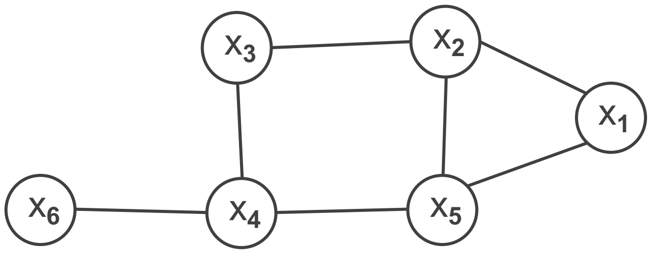

Figure 1: Graphical model for question 2.

iii. $x_1 \perp\!\!\!\perp x_3 | x_2, x_5$

iv. $x_6 \perp\!\!\!\perp x_1 | x_2, x_3, x_4, x_5$

v. $x_6 \perp\!\!\!\perp x_1 | x_2, x_4$

vi. $x_6 \perp\!\!\!\perp x_1 | x_2$

vii. $x_6 \perp\!\!\!\perp x_1 | x_4$

viii. $x_6, x_1 \perp\!\!\!\perp x_3, x_5 | x_2, x_4$

ix. $x_6, x_1 \perp\!\!\!\perp x_3, x_5 | x_4$