Homework #4

Due: Th. Feb. 12

Learning in Graphical Models ICS 280
URL: http://www.ics.uci.edu/~welling/teaching/GraphicalModels.html
Instructor :Max Welling
Office: 414C
phone: 824-8169
email: welling@ics.uci.edu

1. Read handouts provided in class.

2. Identify the sufficient statistics, the canonical parameters, the moment parameters, log-partition function $A(\eta)$ and the mapping $\psi(\cdot)$ between moment and canonical representation for a multi-variate Gaussian distribution $\mathcal{G}_{\mathbf{x}}[\mu;\Sigma]$.

3. a. In a bound optimization algorithm we like the following properties to hold:

$$(A)\ \ Q(\theta,\theta^t) \leq L(\theta^t)\ \ \forall\theta \tag{1}$$

$$(B)\ \ Q(\theta=\theta^t,\theta^t) = L(\theta^t) \tag{2}$$

In class we have proven property A for the GIS algorithm. Now prove property (B) for the bound that was derived in class.

b. The algorithm derived in class was based on Jensen's bound:

$$e^{\sum_i \pi_i z_i} \leq \sum_i \pi_i e^{z_i} \Rightarrow e^{\sum_i \frac{f_i(x)}{f^{\#}(x)} f^{\#}(x)\Delta_i^t} \leq \sum_i \frac{f_i(x)}{f^{\#}(x)} e^{\Delta_i^t f^{\#}(x)} \tag{3}$$

with $\sum_i \pi_i = 1$. However, for $f^{\#} \neq c$ this still lead to one-dimensional non-linear optimization problems. We will now introduce a different bound and derive a new IS algorithm,

$$e^{\sum_i \pi_i z_i} \leq \sum_i \pi_i e^{z_i} + (1 - \sum_i \pi_i) \qquad \sum_i \pi_i \leq 1 \tag{4}$$

Prove the above bound from Jensen's bound by setting a subset of the z_i equal to 0 (hint: the sums $\sum_i$ in Jensen and the generalized bound will range over different domains).

c. Using this new bound, derive a new IS algorithm that has analytical updates even in the case where $\sum_i f_i(x) \neq c$. You may start from the point where:

$$\ell(\theta|D) \geq \sum_i \theta_i \sum_x \tilde{p}(x) f_i(x) - \sum_x p(x|\theta^t) e^{\sum_i \Delta\theta_i^t f_i(x)} - \log Z(\theta^t) + 1 \tag{5}$$

and introduce $f^{max} = \max_x \sum_i f_i(x)$.

d. Show that also in this case the property (B) holds implying we still have a bound optimization algorithm.

e. Although the extra bound simplifies the update rule for the general case when $\sum_i f_i(x) \neq c$ it can sometimes lead to very slow convergence. Can you predict from the features when this will be the case - i.e. what property of the features will result in slow convergence when using the above bound?

4. We consider a Poisson model $p(n) = \lambda^n e^{-\lambda}/n!$ for count data: $n = 0, 1, ...\infty$.

   a. Identify the canonical parameter η, the sufficient statistic $T(n)$, the function $h(n)$, the log-partition function $A(\eta)$ and the mapping $\psi(\cdot)$.

   b. Use $A(\eta)$ to show that the mean count $\mathbf{E}[n] = \lambda$, and the variance $\mathbf{E}[n - \lambda]^2 = \lambda$ (i.e. the variance is equal to the mean).

   c. Assume we have inputs $x_1, x_2, ..., x_d$ (arbitrary domain) and a count label $y \in [0, 1, 2, ...]$. We wish to use a generalized linear model with $p(y|x)$ a Poisson model with conditional mean: $\mathbf{E}[y|x] = f(\theta^T x)$. What is $f(\cdot)$ if we use the canonical link function?

   d. Using the canonical link function, give the updates for the IRLS algorithm. You may simply copy the general update rule from the book and give the relevant expressions that are case dependent - i.e. the weight-matrix and the mean.