# Homework $\#3$

**Due: Th. Feb. 05**

**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. **Iterative Proportional Fitting and Maximum Entropy Models**
   In the following exercise we will look at the IPF algorithm for the following very simple model: 3 observed random variables in a chain interacting through potentials $\psi_{12}(x_1, x_2)$ and $\psi_{23}(x_2, x_3)$. Each RV can be in one out of $D$ states.

   a. Draw the graphical model and write the expression for the probability distribution in terms of the above potential functions.
   b. Is this graph decomposable? Given a data-set $\{x_i\},\ i = 1..N$, give a maximum likelihood assignment for the potential functions and the normalization constant in terms of the data.
   c. We will now run IPF. Initialize the potentials to "all ones" and update the potentials (symbolically) in the order $\psi_{12}$ and $\psi_{23}$ etc. In how many iterations did IPF converge? Compare your answer with that of (b).
   d. We now assume that someone else has received the actual data who, for some reason, does not wish to share this with you. The only information she wishes to transmit is the following 3 empirical marginals: $\tilde{p}_i(x_i) = m_1(x_i)/N,\ i = 1, 2, 3$. Rewrite these marginals as expectations over features of the form $f_{y_i}(x_i) = \delta(y_i, x_i),\ i, 1, 2, 3$, i.e. there are $3D$ different features (treating $y_i$ as a label). What is the general form of the maximum entropy distribution, $p^{MaxEnt}(x_1, x_2, x_3)$, in terms of these features and parameters $\theta_i(y_i),\ i = 1, 2, 3$?
   e. Give the expression for the parameters $\theta_i(y_i),\ i = 1, 2, 3$ in terms of the marginals $\tilde{p}_i(x_i) = m_1(x_i)/N,\ i = 1, 2, 3$, such that the feature-constraints are satisfied.

3. **Proof of Convexity for Fully observed Models** The expression for the log-likelihood of an undirected model in the feature representation is:

$$\ell(\theta) = \sum_{i=1}^{F} \theta_i f_i(x) - \log Z(\theta) \tag{1}$$

   a. Prove the following identities:

$$\mathbf{E}[f_i(x)] \doteq \sum_x p(x) f_i(x) = \frac{\partial}{\partial \theta_i} \log Z(\theta) \tag{2}$$

$$\mathbf{E}[f_i(x) f_j(x)] - \mathbf{E}[f_i(x)]\mathbf{E}[f_j(x)] = \frac{\partial^2}{\partial \theta_i \partial \theta_j} \log Z(\theta) \tag{3}$$

   b. The second identity is the "variance-covariance" matrix $C_{ij}$ between the features. Rewrite this as,

$$C_{ij} = \mathbf{E}[(f_i(x) - \mathbf{E}[f_i(x)])(f_j(x) - \mathbf{E}[f_j(x)])] \tag{4}$$

   c. To show that $C$ is positive definite we must show that for any $\alpha$ with $||\alpha|| = 1$ we have $\alpha^T C \alpha > 0$. Use the expression from (b) to prove positive definiteness of the covariance (hint: use the fact that the expectation of any function $g(x)^2$ must be larger than $0$.

   d. Now show that,

$$\frac{\partial^2 \ell(\theta)}{\partial \theta_i \partial \theta_j} < 0 \tag{5}$$

   which proves that $\ell(\theta)$ is a concave function of $\theta_1, ..., \theta_D$. This implies that there is only one solution where all derivatives vanish, and this solution is a maximum. Thus, IPF must find it.