6.892 Machine learning seminar

http://www.ai.mit.edu/courses/6.892-ml

Prof. Tommi S. Jaakkola
tommi@ai.mit.edu

March 31, 1999

This assignment is due **April 7**. You can either email the assignment to me as plain text (use subject "HOMEWORK 7") or give it to me in class.

1. Assume we have a distribution P over discrete variables $x_1, \ldots, x_n$. Let $\{P_i(x_i)\}$ be the (node) marginals computed from this distribution. Suppose we would like to maximize

$$- \sum_{x_1,\ldots,x_n} P(x_1, \ldots, x_n) E(x_1, \ldots, x_n) + H(P) \tag{1}$$

   with respect to P subject to $P_i(x_i) = \tilde{P}_i(x_i)$ for all i. Show that this is equivalent to first finding the unconstrained max-entropy distribution P^* and consequently using the iterative proportial fitting algorithm to force the marginals of P^* to become $\tilde{P}_i(x_i)$ as required.

2. Formulate the mean field approximation for generic directed graphical models as a dynamical system. Show that for your choice of dynamics the variational lower bound $L(P, x_v)$ serves as a Lyapunov function satisfying

$$\frac{d}{dt} L(P, x_v) \geq 0 \tag{2}$$

   and $\frac{d}{dt} L(P, x_v) = 0$ if and only if $\frac{d}{dt} P_i(x_i) = 0$ for all i and x_i, where $P_i(x_i)$ is a component of the mean field distribution.

3. **Research level**: For graphical models even approximate inference is NP-hard, i.e., roughly speaking it is almost as hard to compute the probabilities of interest within $\epsilon > 0$ accuracy than it is to compute them exactly. It is therefore important to characterize a/the subclass of graphical models for which approximate inference is feasible. Towards answering this question, provide a set of conditions for directed graphical models so that mean field approximation will be accurate. Hint: start with "trivial" conditions such as that all the variables in the graph are "almost" independent.