Homework #2

Due: Th. Jan.29

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 the handouts provided in class.
2. a. Draw the directed acyclic graph corresponding to the distribution

$$p(x) = p(x_5|x_3, x_4)p(x_4|x_2)p(x_3|x_1)p(x_2|x_1)p(x_1) \tag{1}$$

Moralize this graph into an undirected graph. Is the moralized graph triangulated? If not, give a chord-less cycle.

b. Give the reconstituted graph after graph-elimination in the order $[5, 4, 3, 2, 1]$. Is this graph triangulated? If not, give a chord-less cycle.

c. Give all the maximal cliques in the reconstituted graph. What is the tree-width of the original Bayes-net? (Instead of checking all $5!$ possible elimination algorithms give an argument for your answer).

d. Assume that we observe $x_5 = \bar{x}_5$. Run the elimination algorithm (symbolically) in the order $[5, 4, 3, 2, 1]$ in terms of the conditional probability tables of item [a] to obtain $p(x_1|\bar{x}_5)$ E.g. the first step is:

$$m_5(x_3, x_4) = \sum_{x_5} p(x_5|x_3, x_4)\delta(x_5, \bar{x}_5) \tag{2}$$

etc.

e. Change the original directed graph by removing the edge between x_4 and x_5. Give the expression for $p(x)$ in terms of constituent conditional and marginal distributions. Is the graph triangulated? Is the resulting graph a directed tree?.

f. Turn the graph into an undirected graphical model by moralizing. Give expressions for the potentials of the undirected graph in terms of the conditional and marginal probabilities of the directed graph.

g. We now want to compute all marginal probabilities $p_i(x_i), \quad i = 1, ..., 5$. Run belief propagation (symbolically) according to the "message passing protocol" discussed in class (and the book), starting at the leaf nodes inward to the root nodes and outward to the leaf nodes again. For example, the message from x_3 to x_1 is (not the first message send according to the protocol!)

$$m_{3\to1}(x_1) = \sum_{x_3} \psi_(x_1, x_3)m_{5\to3}(x_3) \tag{3}$$

If more than one message can be send at a time you may break the tie at your convenience.

h. Now that you have computed (symbolically) all possible messages in the graph, give the expressions for all the marginal probabilities in terms of these messages. Do these expressions represent the exact marginal distributions?

i. Assume we go back to the original Bayes' net which includes the edge between node 4 and 5. If you run belief propagation on the corresponding moralized graph, will the message update protocol ever stop?

3. a. In class we have looked at the probabilistic model for linear regression. In this exercise we will look at non-linear regression. Assume we we have a data-set of N cases: $\{x_i, y_i\}$ where y_i are the one dimensional targets and x_i are the d-dimensional inputs. Consider a "neural network" that maps inputs to outputs using a function $y = f_w(x)$ where w are the weights of the network. Just like the linear case, we want a Gaussian probability density $p(y|x)$ with mean $E[y|x] = f_w(x)$. Give the full expression for $p(y|x)$.

b. Write the expression for log-likelihood and compute its derivatives with respect to the weights w and the variance σ^2 (the first derivative is a function of $f'_w(x) \doteq \frac{\partial f_w(x)}{\partial w}$). Given these derivatives, how would you maximize the log-likelihood on a computer?

4. We now turn to the "naive Bayes'" model.

a. Assume we are trying to classify between two labels $y = [0, 1]$ and we have precisely two continuous attributes x_1 and x_2. Write the expression for the Naive Bayes' model.

b. We now get an input $x_1 = \bar{x}_1$ and $x_2 = \bar{x}_2$, and we know from our model the following quantities: $p(y = 0) = 0.1$, $p(\bar{x}_1|y = 0) = 1e - 4$, $p(\bar{x}_1|y = 1) = 1e - 5$, $p(\bar{x}_2|y = 0) = 1e - 6$, $p(\bar{x}_2|y = 1) = 1e - 3$. Compute the "log-odds" ratio,

$$\log \left[\frac{p(y = 1|\bar{x}_1, \bar{x}_2)}{p(y = 0|\bar{x}_1, \bar{x}_2)} \right] \tag{4}$$

How would you classify the input $(\bar{x}_1, \bar{x}_2)$?