CS 281A/Stat 241A Homework Assignment 2 (due September 25)

1. **Multivariate Gaussian.**

   In this problem, we consider some issues related to marginal independence and conditional independence in the setting of the multivariate Gaussian distribution. Consider a Gaussian distribution for a vector $X = (X_1, X_2, X_3)^T$. That is, letting $x = (x_1, x_2, x_3)^T$ and $\mu = (\mu_1, \mu_2, \mu_3)^T$, we have:

$$p(x) = \frac{1}{(2\pi)^{3/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\}, \qquad (1)$$

   where Σ is a 3×3 matrix.

   (a) Suppose that we have:

$$\Sigma = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

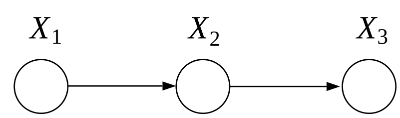
   Are there any marginal independencies among the components of X? What about conditional independencies? (This latter question may be harder to answer and you may want to wait until the end of the problem to answer it).

   (b) Suppose that we have:

$$\Sigma = \begin{bmatrix} 0.75 & 0.5 & 0.25 \\ 0.5 & 1.0 & 0.5 \\ 0.25 & 0.5 & 0.75 \end{bmatrix}$$

   Are any marginal independencies among the components of X? What about conditional independencies? (Again, you may want to wait to answer the latter question).

   (c) Let us now suppose that the distribution on X can be represented by the following graphical model:



   Let $p(x_1)$ be a Gaussian with zero mean and unit variance, let $p(x_2 \mid x_1)$ be a Gaussian with mean x_1 and unit variance, and let $p(x_3 \mid x_2)$ be a Gaussian with mean x_2 and unit variance. Take the product of these three distributions to obtain the joint probability. Expressing this in the general form in (1), read off the inverse covariance matrix Σ^{-1}. Interpret the zeros in this matrix in terms of the conditional independencies encoded in the graph.

   (d) Use Matlab or Splus to compute the inverse of Σ^{-1}, i.e., to compute Σ, in the previous example. Now return to part (b) and answer the question regarding conditional independence. Similarly, return to part (a) and finish the question.

(e) (Bonus). Give a general theorem regarding the relationship between conditional independence for multivariate Gaussian distributions and zeros in the inverse covariance matrix. Prove your theorem.

2. **Naive Bayes**

   Consider a binary classification problem in which a binary label $Y \in \{0,1\}$ is to be predicted from an m-vector $X = (X_1, X_2, \ldots, X_m)$, where $X_i \in \{0,1\}$ for each i.

   Consider a *naive Bayes* model, in which the components X_i are assumed mutually conditionally independent given the class label Y.

   (a) Draw a directed graphical model corresponding to the naive Bayes model.

   (b) Find a mathematical expression for the *posterior class probability* $p(Y = 1 \mid x)$, in terms of the *prior class probability* $p(Y = 1)$ and the *class-conditional densities* $p(x_i \mid y)$.

3. **Logistic and softmax functions**

   (a) The function

   $$y = \frac{1}{1 + e^{-x}}$$

   is known as the *logistic function*. Sketch the graph of the function. Find the derivative dy/dx expressing your result in terms of y rather than x. Sketch the graph of the derivative.

   (b) The function

   $$y_i = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}},$$

   where $i = 1, \ldots, n$ is a generalization of the logistic function known as the *Potts' distribution* or the *softmax function*. Find the derivatives $\partial y_i / \partial x_j$, expressing your result in terms of y_i and y_j. The expression is particularly neat if you use the Kronecker delta function.

4. **The polytree algorithm**

   Consider the following algorithm due to Kim and Pearl for computing singleton marginal probabilities on a general polytree:

   $$\begin{aligned} p(x_i) &\propto \lambda(x_i)\rho(x_i) \\ \lambda(x_i) &= \prod_{j \in \xi_i} \lambda_{ji}(x_i) \end{aligned}$$

$$
\begin{aligned}
\rho(x_i) &= \sum_{x_{\pi_i}} p(x_i \mid x_{\pi_i}) \prod_{j \in \pi_i} \rho_{ji}(x_j) \\
\lambda_{ji}(x_i) &= \sum_{x_j} \lambda(x_j) \sum_{x_{\pi_j} \setminus i} p(x_j \mid x_{\pi_j}) \prod_{k \in \pi_j \setminus i} \rho_{kj}(x_k) \\
\rho_{ji}(x_j) &= \prod_{k \in \xi_j \setminus i} \lambda_{kj}(x_j) \sum_{x_{\pi_j}} p(x_j \mid x_{\pi_j}) \prod_{l \in \pi_j} \rho_{lj}(x_l),
\end{aligned}
$$

where π_i denotes the set of parents of node i and ξ_i denotes the set of children of node i. Note that $\lambda_{ji}(x_i)$ is the message sent to node i from its jth child, and $\rho_{ji}(x_j)$ is the message sent to node i from its jth parent.

(a) Draw a generic polytree as a factor graph.

(b) Derive the Kim-Pearl polytree algorithm by starting from the sum-product algorithm for factor graphs.

5. **BONUS**

Prove that the sum-product algorithm correctly computes singleton marginal probabilities for all nodes in an undirected tree.