290 C

S04

HW 2

Handed out:
F, 4-23-04

Due
Tu, 5-4-04
in class

1. (SUM-PRODUCT **Algorithm)**

(a) Consider a sequence of random variables $X_1, \ldots, X_T$ that lie in the set $\{1, 2, 3\}$. Suppose that $P(X_1 = 1) = 1/2$, and $P(X_1 = 2) = P(X_1 = 3) = 1/4$. Suppose also that given X_{t-1}, the distribution of X_t is conditionally independent of X_s at all earlier times, $s < t-1$. At each time $t \in \{1, \ldots, T\}$, the random variable Y_t is the indicator function of $X_t = 3$, that is, $Y_t \in \{0, 1\}$ and $Y_t = 1$ iff $X_t = 3$. Specify a directed graphical model for the set of random variables $\{X_1, \ldots, X_T, Y_1, \ldots, Y_T\}$.

(b) For the model of (1a), suppose that the conditional distribution of X_t given X_{t-1} is identical for all $t > 1$, and the matrix of transition probabilities (with entries $A_{ij} = P(X_{t+1} = j | X_t = i)$) is

$$A = \frac{1}{6} \begin{bmatrix} 3 & 2 & 1 \\ 2 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}.$$

Suppose that we observe $\bar{y}_1 = 0, \bar{y}_2 = 0, \bar{y}_3 = 1$, and we wish to use the SUM-PRODUCT algorithm to compute $p(x_t | \bar{y}_1, \bar{y}_2, \bar{y}_3)$ for $t = 1, 2, 3$. What messages will be passed by the algorithm? What will the algorithm return?

(You'll find it convenient to use matrix notation and to write messages as vectors:

$$m_{y_t, x_t} = [m_{y_t, x_t}(1), m_{y_t, x_t}(2), m_{y_t, x_t}(3)].$$

You might use matlab or a similar language.)

(c) Suppose that we augment the model of (1b) as follows. Let Z_t be a Gaussian random variable with mean Y_t and unit variance, and suppose that given Y_t, Z_t is conditionally independent of all other variables. Specify a directed graphical model for the set of random variables $\{X_1, \ldots, X_T, Y_1, \ldots, Y_T, Z_1, \ldots, Z_T\}$.

(d) Suppose that $z_1 = 0.5$, $z_2 = -0.5$, $z_3 = 2.0$. Compute $p(x_i | z_1, z_2, z_3)$ for $i = 1, \ldots, 3$ under the assumptions of (1c).

(e) Consider the model of (1a), but with the transition probability matrix

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Specify a simpler graphical model for this case. What messages will be passed by the SUM-PRODUCT algorithm in this case if we wish to compute $p(x_t | \bar{y}_1, \bar{y}_2, \bar{y}_3)$? Explain how these messages are equivalent to the messages that would be passed in the full graphical model of (1a).

(f) Consider the model of (1a), but with the transition probability matrix

$$A = \frac{1}{4} \begin{bmatrix} 2 & 1 & 1 \\ 2 & 1 & 1 \\ 2 & 1 & 1 \end{bmatrix}.$$

Specify a simpler graphical model for this case. What messages will be passed by the SUM-PRODUCT algorithm in this case if we wish to compute $p(x_t | \bar{y}_1, \bar{y}_2, \bar{y}_3)$? Explain how these messages are equivalent to the messages that would be passed in the full graphical model of (1a).

3. **(Naive Bayes)**
In a pattern classification problem, a binary label $Y \in \{0,1\}$ is to be predicted from the covariates $X_1, \ldots, X_d \in \{0,1\}$. A *naive Bayes* model assumes that, given the class label Y, the components X_i are conditionally independent.

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

(b) Express the posterior class probability, $p(Y=1|x)$, in terms of the prior class probability $p(Y=1)$ and the class conditionals, $p(x_i|y)$.

(c) Suppose we wish to use a naive Bayes to classify web pages into two classes, and let each X_w be the indicator function of the presence of word w on the page. Explain why this might not be an accurate model of the joint distribution.

(d) Suppose we wish to make a prediction $\hat{y} \in \{0,1\}$. It is easy to show that predicting $\hat{y}=1$ iff $p(Y=1|x) \geq 1/2$ minimizes $p(Y \neq \hat{y})$. Show that making this prediction using the posterior class probability for a naive Bayes model corresponds to a linear classifier, for which $\hat{y}=1$ iff

$$\sum_{i=1}^{d} a_i X_i \geq b$$

for some real numbers $a_1, \ldots, a_d, b$.