Statistical Methods for Artificial Intelligence, Autumn 2006
Problem Set 7, not to be turned in
(Practice Problem for the Final Exam)

Problem 1. This problem is on EM. Suppose that we want to classify $x \in \mathcal{X}$ into one of K classes using a single feature vector $\Phi(x)$ where each feature is either 0 or 1, i.e., for each feature i and $x \in \mathcal{X}$ we have $\Phi_i(x) \in \{0, 1\}$. The naive Bayes model is a generative model for generating pairs $\langle x, y \rangle$ with $x \in \mathcal{X}$ and y a class label, i.e., $y \in \{1, \ldots, K\}$. The naive Bayes model can be defined by the following equations.

$$\begin{aligned} \beta_j &= P(y = j) \\ \beta_{i,j} &= P(\Phi_i(x) = 1 \mid y = j) \end{aligned}$$

$$P_\beta(x, y) = \beta_y \prod_i \begin{cases} \beta_{i,y} & \text{if } \Phi_i(x) = 1 \\ (1 - \beta_{i,y}) & \text{if } \Phi_i(x) = 0 \end{cases}$$

Suppose that we have a sample S of unlabeled values x_1, …, x_T from $\mathcal{X}$.

a. Give equations for $P_\beta(y_t = j|x_t)$.

b. Now suppose we want the maximum likelihood (ML) value of β under the naive Bayes model.

$$\beta^* = \underset{\beta}{\operatorname{argmax}} \prod_{t=1}^{T} P_\beta(x_t)$$

The EM algorithm takes some initial vector β^0 and computes values β^1, β^2, β^3, … which improves the above objective. Give equations specifying β^{n+1} as a function of β^n.