CSE 250A Quiz 5

Tuesday November 6, 2012

Instructions. You should do this quiz in partnership with exactly one other student. Write both your names at the top of this page. Discuss the answer to the question with each other, and then write your joint answer below the question. It is ok if you overhear other students' discussions, because you still need to decide if they are right or wrong. You have seven minutes.

Consider the Bayesian network $W \rightarrow Z \rightarrow W'$ where the random variables W and W' are observed, while Z is hidden. Let the training data be $\{(w_t, w'_t)\}_{t=1}^n$. The M step of EM for the edge $W \rightarrow Z$ is the update

$$p(Z = z|W = w) \quad := \quad \frac{\sum_{t=1}^n p(Z = z, W = w|W = w_t, W' = w'_t)}{\sum_{t=1}^n p(W = w|W = w_t, W' = w'_t)}.$$

Simplify the expression on the right as much as possible. Use an indicator function and a count function.

Answer. Change notation from $W = w$ to $W = v$ on both sides, to minimize confusion. A simplification that specifically uses one indicator function and one count function is

$$\begin{aligned} p(Z = z|W = v) \quad &:= \quad \frac{\sum_{t=1}^n p(W = v|Z = z, W = w_t, W' = w'_t)p(Z = z|W = w_t, W' = w'_t)}{\sum_{t=1}^n p(W = v|W = w_t)} \\ &= \quad \frac{\sum_t I(v = w_t)p(Z = z|W = w_t, W' = w'_t)}{\sum_t I(v = w_t)} \\ &= \quad \frac{\sum_t I(w_t = v)p(Z = z|W = w_t, W' = w'_t)}{\text{count}(v = w_t)}. \end{aligned}$$

Additional comments. In the numerator, $p(Z = z|W = w_t, W' = w'_t)$ cannot be simplified further, because both W and W' provide information about the value of Z. Evaluating this probability using Bayes' rule is the E step of EM.

The simplified result is intuitively reasonable. The indicator functions say that the estimated probability that $Z = z$ given $W = w$ depends only on training examples which have this particular w value. The numerator says that this target probability should be high if z is compatible with the observed w_t and w'_t values.