CS613 \ Assignment #4 Solutions

Ariel Stolerman

1)

Let there be a density model with the mixture distribution $p(x) = \sum_{k=1}^{K} \pi_k p(x|k)$, and x is partitioned into 2 parts $x = (x_a, x_b)$. Following is a proof that the conditional density $p(x_b|x_a)$ is itself a mixture distribution, with expressions for the mixing coefficients and the component densities.

We need to show that:

$$p(x_b|x_a) = \sum_{k=1}^{K} \lambda_k p(x_b|x_a, k)$$

The marginal density of x_a is:

$$p(x_a) = \sum_{k=1}^{K} \pi_k p(x_a|k)$$

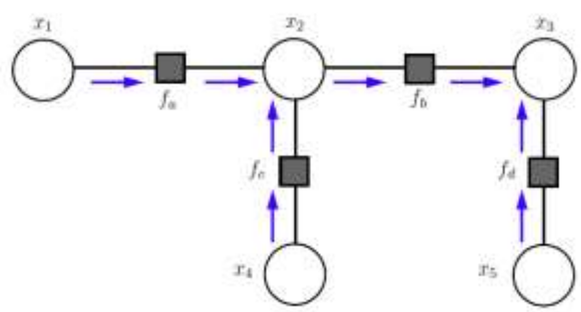
…

2)

2.1) The unnormalized joint distribution of the factor graph is:

$$\tilde{p}(\boldsymbol{x}) = f_a(x_1, x_2) f_b(x_2, x_3) f_c(x_2, x_4) f_d(x_3, x_5)$$

2.2) The marginal distribution of x_3 by the sum-product algorithm is:



$\mu_{x_1 \to f_a}(x_1) = 1$

$$\mu_{f_a \to x_2}(x_2) = \sum_{x_1} f_a(x_1, x_2)$$

$\mu_{x_4 \to f_c}(x_4) = 1$

$$\mu_{f_c \to x_2}(x_2) = \sum_{x_4} f_c(x_2, x_4)$$

$\mu_{x_2 \to f_b}(x_2) = \mu_{f_a \to x_2}(x_2) \mu_{f_c \to x_2}(x_2)$

$$\mu_{f_b \to x_3}(x_3) = \sum_{x_2} f_b(x_2, x_3) \mu_{x_2 \to f_b}(x_2)$$

$\mu_{x_5 \to f_d}(x_5) = 1$

$$\mu_{f_d \to x_3}(x_3) = \sum_{x_5} f_d(x_3, x_5)$$

⇒ The marginal distribution of x_3 is: $\mu_{f_b \to x_3}(x_3) \mu_{f_d \to x_3}(x_3) = \boxed{\sum_{x_2} f_b(x_2, x_3) \sum_{x_1} f_a(x_1, x_2) \sum_{x_4} f_c(x_2, x_4) \sum_{x_5} f_d(x_3, x_5)}$

3)

Following is a proof that if any elements of the parameters π or A for a hidden Markov model are initially set to zero, then those elements will remain zero in all subsequent updates of the EM algorithm.

We know the update rules for the starting probability π and the transition probability matrix A are:

$$\pi_k = \frac{\gamma(z_{1k})}{\sum_{j=1}^{K} \gamma\left(z_{1j}\right)}, \quad A_{jk} = \frac{\sum_{n=2}^{N} \xi\left(z_{n-1,j}, z_{nk}\right)}{\sum_{l=1}^{K} \sum_{n=2}^{N} \xi\left(z_{n-1,j}, z_{nl}\right)}$$

In the E step there will be no training samples associated with 0-probability states or transitions to 0-probability transitions. Therefore the updated probabilities in the Mstep will remain 0 as well.

Looking at the joint posterior of two successive latent variables, $\xi(z_{n-1}, z_n)$:

$$\xi(z_{n-1}, z_n) = p(z_{n-1}, z_n | X) = \frac{p(X|z_{n-1}, z_n) p(z_{n-1}, z_n)}{p(X)} =$$

$$\frac{p(x_1, \dots, x_{n-1}|z_{n-1}) p(x_n|z_n) p(x_{n+1}, \dots, x_N|z_n) p(z_n|z_{n-1}) p(z_{n-1})}{p(X)} \underset{\substack{A_{jk}=0 \Rightarrow \\ p(z_n|z_{n-1})=0}}{=} 0$$

Therefore all future updates are zero.