STA 414/2104, Spring 2014, Assignment #2 — Derivation of M step

The assignment handout says the M step update for $\theta_{k,j}$ should be

$$\hat{\theta}_{k,j} \;=\; \frac{\alpha + \sum_{i=1}^n r_{i,k} x_{i,j}}{2\alpha + \sum_{i=1}^n r_{i,k}}$$

where $r_{i,k}$ is the probability that case i came from component k, estimated in the E step.

The general form of the M step is to maximize $E_Q[\log P(x,z|\theta) + G(\theta)]$, where $G(\theta)$ is the "penalty". (We can ignore the π parameters here, since they don't interact with θ, given that the distribution Q found in the E step is fixed during the M step.) All we really need to know about Q is the responsibilities, $r_{i,k}$, found in the E step as the expectations of the $z_{i,k}$, which indicate which mixture component each data item came from (as in the week 7 lecture slides).

We can write the terms of $E_Q[\log P(x,z|\theta) + G(\theta)]$ that involve $\theta_{k,j}$ as

$$E_Q\Big[\log\Big(\prod_{i=1}^n \Big(\theta_{k,j}^{x_{i,j}}(1-\theta_{k,j})^{1-x_{i,j}}\Big)^{z_{i,k}}\Big) + \alpha(\log(\theta_{k,j}) + \log(1-\theta_{k,j}))\Big]$$

where note that each $z_{i,k}$ will be either zero one. This can be rewritten as

$$\begin{aligned}
&E_Q\Big[\sum_{i=1}^n z_{i,k}\log\Big(\theta_{k,j}^{x_{i,j}}(1-\theta_{k,j})^{1-x_{i,j}}\Big) + \alpha(\log(\theta_{k,j}) + \log(1-\theta_{k,j}))\Big]\\
&\quad=\; \sum_{i=1}^n r_{i,k}\log\Big(\theta_{k,j}^{x_{i,j}}(1-\theta_{k,j})^{1-x_{i,j}}\Big) + \alpha(\log(\theta_{k,j}) + \log(1-\theta_{k,j}))\\
&\quad=\; \sum_{i=1}^n \Big(r_{i,k}x_{i,j}\log(\theta_{k,j})\Big) + \sum_{i=1}^n \Big(r_{i,k}(1-x_{i,j})\log(1-\theta_{k,j})\Big) + \alpha(\log(\theta_{k,j}) + \log(1-\theta_{k,j}))\\
&\quad=\; \Big(\alpha + \sum_{i=1}^n r_{i,k}x_{i,j}\Big)\log(\theta_{k,j}) + \Big(\alpha + \sum_{i=1}^n r_{i,k}(1-x_{i,j})\Big)\log(1-\theta_{k,j})
\end{aligned}$$

To find the maximum with respect to $\theta_{k,j}$, we take the derivative and set it equal to zero:

$$\Big(\alpha + \sum_{i=1}^n r_{i,k}x_{i,j}\Big)\Big/\theta_{k,j} \;-\; \Big(\alpha + \sum_{i=1}^n r_{i,k}(1-x_{i,j})\Big)\Big/(1-\theta_{k,j}) \;=\; 0$$

Multiply this equation by $\theta_{k,j}(1-\theta_{k,j})$, we get

$$\Big(\alpha + \sum_{i=1}^n r_{i,k}x_{i,j}\Big)(1-\theta_{k,j}) \;-\; \Big(\alpha + \sum_{i=1}^n r_{i,k}(1-x_{i,j})\Big)\theta_{k,j} \;=\; 0$$

which gives

$$\alpha + \sum_{i=1}^n r_{i,k}x_{i,j} \;=\; \Big(\alpha + \sum_{i=1}^n r_{i,k}x_{i,j}\Big)\theta_{k,j} \;+\; \Big(\alpha + \sum_{i=1}^n r_{i,k}(1-x_{i,j})\Big)\theta_{k,j}$$

One can now easily see that the solution to this is the update above.