EM Supplment

The EM algorithm is often cited for decreasing the observed data log likelihood at each step. A proof of this result is outline below.

Define the observed data, $\mathbf{y} = (y_1, \ldots, y_n)'$, and the unobserved data, $\mathbf{z} = (z_1, \ldots, z_n)'$, using the same notation as we did in class. Next we write the complete data likelihood as the product of the observed data likelihood and a conditional term:

$$g(\mathbf{y}, \mathbf{z}|\psi) = g(\mathbf{y}|\psi)g(\mathbf{z}|\mathbf{y}, \psi),$$

where ψ is the parameter vector. In the normal mixture context, ψ would be comprised of the elements of π, μ, and $\sigma^{\mathbf{2}}$. Take the natural log of both sides, rearrange terms, and simplify the notation to obtain the complete data log likelihood $l(\psi|\mathbf{y}, \mathbf{z})$ and observed data log likelihood $l(\psi|\mathbf{y})$

$$\ln g(\mathbf{y}, \mathbf{z}|\psi) = \ln g(\mathbf{y}|\psi) + \ln g(\mathbf{z}|\mathbf{y}, \psi)$$

$$l(\psi|\mathbf{y}) = l(\psi|\mathbf{y}, \mathbf{z}) - \ln g(\mathbf{z}|\mathbf{y}, \psi). \tag{1}$$

We now take the expectation of both sides of (1) with respect to the density $g(\mathbf{z}|\mathbf{y}, \psi^{(\mathbf{m})})$; this is the E step. Note that $l(\psi|\mathbf{y})$ does not depend on $\mathbf{z}$ and so is unchanged. The first term on the right hand side will be the usual term $Q(\psi, \psi^{(\mathbf{m})})$ that we maximize in the M step.

$$l(\psi|\mathbf{y}) = Q(\psi, \psi^{(\mathbf{m})}) - \int \ln g(\mathbf{z}|\mathbf{y}, \psi)g(\mathbf{z}|\mathbf{y}, \psi^{(\mathbf{m})})d\mathbf{z}.$$

$$l(\psi|\mathbf{y}) = Q(\psi, \psi^{(\mathbf{m})}) - H(\psi, \psi^{(\mathbf{m})})$$

We will next look at the difference in $l(\mathbf{psi}|\mathbf{y})$ after a given EM step. We can show that the first term in the difference is nonnegative, and the second term is nonpositive, so that on the whole, the EM step increases the observed data log likelihood.

$$l(\psi^{(\mathbf{m+1})}|\mathbf{y}) - l(\psi^{(\mathbf{m})}|\mathbf{y}) =$$

$$\left(Q(\psi^{(\mathbf{m+1})}, \psi^{(\mathbf{m})}) - Q(\psi^{(\mathbf{m})}, \psi^{(\mathbf{m})})\right) -$$

$$\left(H(\psi^{(\mathbf{m+1})}, \psi^{(\mathbf{m})}) - H(\psi^{(\mathbf{m})}, \psi^{(\mathbf{m})})\right).$$

Note that the first difference on the right-hand side is greater than or equal to 0, since $\psi^{(\mathbf{m+1})}$ maximizes $Q(\psi, \psi^{(\mathbf{m})})$. The second difference can be rearranged:

$$H(\psi^{(\mathbf{m+1})}, \psi^{(\mathbf{m})}) - H(\psi^{(\mathbf{m})}, \psi^{(\mathbf{m})}) =$$

$$E_{\mathbf{z}|\mathbf{y},\psi^{(\mathbf{m})}}\left[\ln g(\mathbf{z}|\mathbf{y},\psi^{(\mathbf{m+1})}) - \ln g(\mathbf{z}|\mathbf{y},\psi^{(\mathbf{m})})\right] =$$

$$E_{\mathbf{z}|\mathbf{y},\psi^{(\mathbf{m})}}\left[\ln \frac{g(\mathbf{z}|\mathbf{y},\psi^{(\mathbf{m+1})})}{g(\mathbf{z}|\mathbf{y},\psi^{(\mathbf{m})})}\right].$$

At this point, we can apply Jensen's inequality and simplify terms to get our result:

$$E_{\mathbf{z}|\mathbf{y},\psi^{(\mathbf{m})}}\left[\ln \frac{g(\mathbf{z}|\mathbf{y},\psi^{(\mathbf{m+1})})}{g(\mathbf{z}|\mathbf{y},\psi^{(\mathbf{m})})}\right] \leq$$

$$\ln\left[E_{\mathbf{z}|\mathbf{y},\psi^{(\mathbf{m})}}\left[\frac{g(\mathbf{z}|\mathbf{y},\psi^{(\mathbf{m+1})})}{g(\mathbf{z}|\mathbf{y},\psi^{(\mathbf{m})})}\right]\right] =$$

$$ln \int g(\mathbf{z}|\mathbf{y},\psi^{(\mathbf{m+1})})d\mathbf{z} = \ln(1) = 0.$$

So the second difference is less than or equal to 0, which implies $l(\psi^{(\mathbf{m+1})}|\mathbf{y}) - l(\psi^{(\mathbf{m})}|\mathbf{y}) \geq 0$ and the result is proven.