Gaussian Mixture Model by Using EM

Shubing Wang*

April 17, 2004

1. *2-component Gaussian Mixture Model.* We consider the case that all the 5 parameters

$$\theta^T = (p, \mu_1, \sigma_1, \mu_2, \sigma_2)$$

are unknown. By using EM algorithm:

- **Estimation Step:** define $\tilde{p_i} = P(y_i \sim f_1|\Theta)$, then

$$\tilde{p} = \frac{p f_1(y_i|\Theta)}{p f_1(y_i|\Theta) + (1-p) f_2(y_i|\Theta)}.$$

So

$$\begin{aligned} Q(\Theta, \Theta') &= \sum_i \{[\tilde{p_i}(\log p' - \log\sqrt{2\pi} \\ &\quad - \log\sigma_1' - \frac{(Y_i - \mu_1')^2}{2\sigma_1^2})] \\ &\quad + [(1-\tilde{p_i})(\log(1-p') - \log\sqrt{2\pi} \\ &\quad - \log\sigma_2' - \frac{(Y_i - \mu_2')^2}{2\sigma_2^2})]\} \end{aligned}$$

- **Maximization Step:**

$$\begin{aligned} p' &= \frac{\sum_i \tilde{p_i}}{n} \\ \mu_1' &= \frac{\sum_i \tilde{p_i} Y_i}{\sum_i \tilde{p_i}} \\ \sigma_1' &= \sqrt{\frac{\sum_i \tilde{p_i}(Y_i - \mu_1')^2}{\sum_i \tilde{p_i}}} \\ \mu_2' &= \frac{\sum_i (1-\tilde{p_i}) Y_i}{\sum_i (1-\tilde{p_i})} \\ \sigma_2' &= \sqrt{\frac{\sum_i (1-\tilde{p_i})(Y_i - \mu_2')^2}{\sum_i (1-\tilde{p_i})}} \end{aligned}$$

2. *M-component Gaussian Mixture Model.* It can derived similarly by using EM with the Maximization step:

$$\begin{aligned} p_j' &= \frac{\sum_i \tilde{p_{ji}}}{n}, \qquad 1 \le j \le M \\ \mu_j' &= \frac{\sum_i \tilde{p_{ji}} Y_i}{\sum_i \tilde{p_{ji}}}, \quad \text{if} \quad 1 \le j < M \\ \sigma_j' &= \sqrt{\frac{\sum_i \tilde{p_{ji}}(Y_i - \mu_j')^2}{\sum_i \tilde{p_{ji}}}}, \quad \text{if} \quad 1 \le j < M \\ \mu_M' &= \frac{\sum_i (1 - \sum_{j=1}^{M-1} \tilde{p_{ji}}) Y_i}{\sum_i (1 - \sum_{j=1}^{M-1} \tilde{p_{ji}})} \\ \sigma_M' &= \sqrt{\frac{\sum_i (1 - \sum_{j=1}^{M-1} \tilde{p_{ji}})(Y_i - \mu_M')^2}{\sum_i (1 - \sum_{j=1}^{M-1} \tilde{p_{ji}})}} \end{aligned}$$

*Email: wang6@wisc.edu

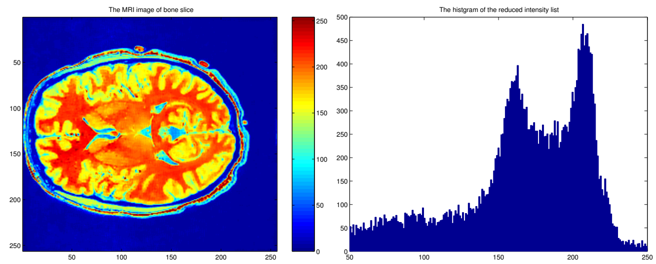

Figure 1: Left: mid sagittal brain image. Right: histogram of image intensity.

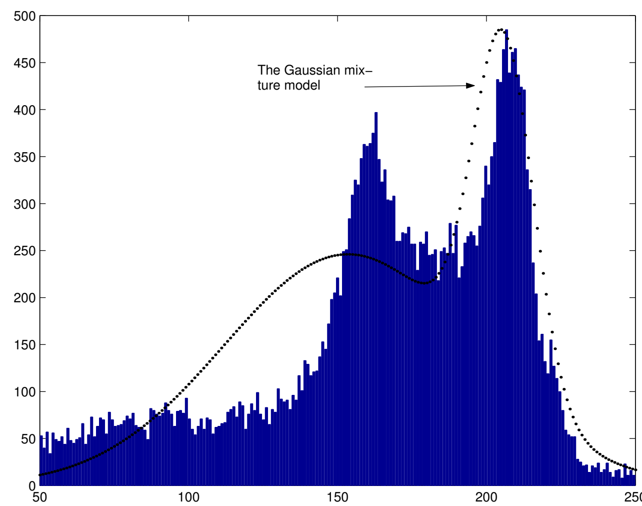

Figure 2: The comparation of the original data and the fitted Gaussian Mixture Model

3. *Segmentation by Using Gaussian Mixture Model.* We start with data as shown in Figure 1. We always should get the histgram as well, since the initial guess is crucial when one is trying to do EM algorithm. A good initial guess will absolutely improve the preformance of the programming and histgrams provide the clue for goog initial guess. One can get the 2-component Gaussian Mixture Model as shown in Figure 2. The result of segmentation as shown in Figure 3.

   Intuitively, 3-component Gaussian Mixture Model will give a better result. Figure 4 and Figure 5 show the result of segmentation by using the 3-component Gaussian Mixture Model.

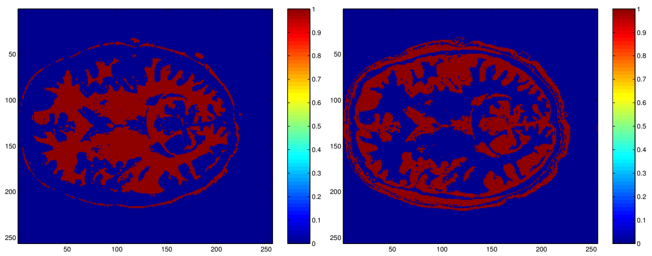

Figure 3: The two components of the 2-component Gaussian Mixture Model

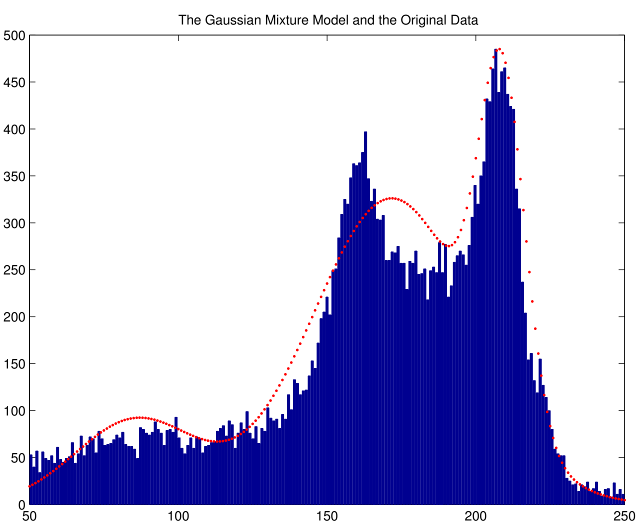

Figure 4: The fitting 3-component Gaussian Mixture Model

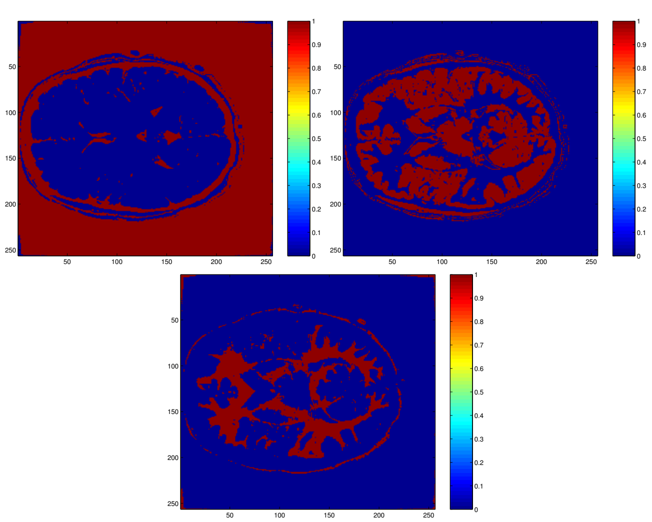

Figure 5: The 3 segments of the 3-component Gaussian Mixture Model