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

\[ \theta^1 = (\rho, \mu_1, \sigma_1, \mu_2, \sigma_2) \]

are unknown. By using EM algorithm:

- **Estimation Step:** define \( \bar{p}_i = P(y_i \sim f_1(\Theta)) \), then
  \[
  \hat{p} = \frac{p f_1(y_i|\Theta)}{p f_1(y_i|\Theta) + (1 - \hat{p}) f_2(y_i|\Theta)}.
  \]
  So
  \[
  Q(\Theta, \Theta') = \sum_i \{ | \bar{p}_i | \log \hat{p} - \log 2 \pi \\
  - \log \sigma_1' - \frac{(Y_i - \mu_1')^2}{2\sigma_1'^2} \}
  + \{ (1 - \bar{p}_i) \log (1 - \hat{p}) - \log \sqrt{2\pi} \\
  - \log \sigma_2' - \frac{(Y_i - \mu_2')^2}{2\sigma_2'^2} \}
  \]

- **Maximization Step:**
  \[
  \rho' = \frac{\sum_i \bar{p}_i}{n}, \quad 1 \leq j \leq M
  \]
  \[
  \mu_1' = \frac{\sum_i \bar{p}_i Y_i}{\sum_i \bar{p}_i Y_i}, \quad \text{if} \quad 1 \leq j < M
  \]
  \[
  \sigma_1' = \sqrt{\frac{\sum_i \bar{p}_i (Y_i - \mu_1')^2}{\sum_i \bar{p}_i}}
  \]
  \[
  \mu_2' = \frac{\sum_i (1 - \bar{p}_i) Y_i}{\sum_i (1 - \bar{p}_i)}, \quad \text{if} \quad 1 \leq j < M
  \]
  \[
  \sigma_2' = \sqrt{\frac{\sum_i (1 - \bar{p}_i) (Y_i - \mu_2')^2}{\sum_i (1 - \bar{p}_i)}}
  \]

2. M-component Gaussian Mixture Model. It can be derived similarly by using EM with the Maximization step:
  \[
  \rho_j' = \frac{\sum_i \bar{p}_jY_i}{n}, \quad 1 \leq j \leq M
  \]
  \[
  \mu_j' = \frac{\sum_i \bar{p}_jY_i}{\sum_i \bar{p}_j}, \quad \text{if} \quad 1 \leq j < M
  \]
  \[
  \sigma_j' = \sqrt{\frac{\sum_i \bar{p}_j (Y_i - \mu_j')^2}{\sum_i \bar{p}_j}} \]
  \[
  \mu_M' = \frac{\sum_i (1 - \sum_{j=1}^{M-1} \bar{p}_j) Y_i}{\sum_i (1 - \sum_{j=1}^{M-1} \bar{p}_j)}, \quad \text{if} \quad 1 \leq j < M
  \]
  \[
  \sigma_M' = \sqrt{\frac{\sum_i (1 - \sum_{j=1}^{M-1} \bar{p}_j) (Y_i - \mu_M')^2}{\sum_i (1 - \sum_{j=1}^{M-1} \bar{p}_j)}}
  \]

3. Segmentation by Using Gaussian Mixture Model. We start with data as shown in Figure 1. We always should get the histogram 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 histograms provide the clue for good initial guess. One can get the 2-component Gaussian Mixture Model as shown in Figure 2. The result of segmentation as shown in Figure 5.

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 Model.

*Email: wang6@wisc.edu
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