

Stat 471: Lecture 32

Gaussian Mixture and EM algorithm

Moo K. Chung
mchung@stat.wisc.edu

December 1, 2003

1. The image intensity can be modelled as a Gaussian mixture of the form

$$f(y) = pf_1(y) + (1 - p)f_2(y)$$

where $f_1 \sim N(\mu_1, \sigma_1^2)$ and $f_2 \sim N(\mu_2, \sigma_2^2)$. The likelihood function is

$$f(p|y) = \prod_{i=1}^n [pf_1(y_i) + qf_2(y_i)]$$

where $q = 1 - p$. The loglikelihood is

$$L(p|y) = \sum_{i=1}^n \log[pf_1(y_i) + qf_2(y_i)].$$

Solving

$$\frac{\partial L(p|y)}{\partial p} = 0$$

to maximize the loglikelihood is complicated. So we argument data with missing data and apply the EM algorithm.

2. Let X be a Bernoulli random variable with $P(X = 1) = p$. Let $Y \sim f_1$ if $X = 1$ and $Y \sim f_2$ if $X = 0$. Then the joint density $f(x, y)$ of (X, Y) is $f(1, y) = pf_1(y)$ and $f(0, y) = qf_2(y)$. This can be written simply as

$$f(x, y) = [pf_1(y)]^x [qf_2(y)]^{1-x}.$$

The marginal density of Y is obviously

$$f(y) = pf_1(y) + qf_2(y).$$

The conditional density of X given Y is then

$$f(x|y) = \frac{[pf_1(y)]^x [qf_2(y)]^{1-x}}{pf_1(y) + qf_2(y)}.$$

Trivially the conditional expectation

$$\mathbb{E}(X|y, p) = \frac{pf_1(y)}{pf_1(y) + qf_2(y)}. \quad (1)$$

For each observation Y_i , we argument it with missing data X_i which is Bernoulli with the above property. The likelihood for the complete data (x, y) is

$$f(p|x, y) = \prod_{i=1}^n [pf_1(y_i)]^{x_i} [qf_2(y_i)]^{1-x_i}$$

So the loglikelihood for the complete data is

$$\begin{aligned} L(p|x, y) \\ = \sum_{i=1}^n x_i \log[pf_1(y_i)] + (1 - x_i) \log[qf_2(y_i)] \end{aligned}$$

3. Now we take expectation and get Q function

$$\begin{aligned} Q(p|p_0, y) \\ = \mathbb{E}[\log L(p|X, Y)|y, p_0] \\ = \sum_{i=1}^n \mathbb{E}(X_i|y, p_0) \log\left[\frac{pf_1(y_i)}{qf_2(y_i)}\right] + \log[qf_2(y_i)] \end{aligned}$$

From equation (1),

$$\mathbb{E}(X_i|y, p_0) = \frac{p_0 f_1(y_i)}{p_0 f_1(y_i) + q_0 f_2(y_i)}.$$

Neglecting parts that do not contain p ,

$$Q(p|p_0, y) = \sum_{i=1}^n \frac{p_0 f_1(y_i)}{p_0 f_1(y_i) + q_0 f_2(y_i)} \log \frac{p}{1-p} + n \log(1-p).$$

Maximizing Q by $\partial Q / \partial p = 0$, we get

$$p = \frac{1}{n} \sum_{i=1}^n \frac{p_0 f_1(y_i)}{p_0 f_1(y_i) + q_0 f_2(y_i)}.$$

Based on this, we set up iteration

$$\hat{p}_{j+1} = \frac{1}{n} \sum_{i=1}^n \frac{\hat{p}_j f_1(y_i)}{\hat{p}_j f_1(y_i) + (1 - \hat{p}_j) f_2(y_i)}$$

with any initial estimate $\hat{p}_0$. Setting $\hat{p}_0 = 0.5$ will speed up the convergence. Cool?