Stat 992: Lecture 32
Gaussian mixture model II.

Moo K. Chung mchung@stat.wisc.edu

December 14, 2003

Problem. Gaussian mixture

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

where $f_1 \sim N(\mu_1, \sigma_1^2)$ and $f_2 \sim N(\mu_2, \sigma_2^2)$. Assume parameters $\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ are known. We will estimate parameter p by maximizing the likelihood function using the EM algorithm.

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 latent data and apply the EM algorithm.

Solution. 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 of (X, Y) is

$$f(1, y) = pf_1(y), f(0, y) = qf_2(y).$$

This can be written compactly 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 probability of X given Y is then

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

Note $f(x|y)$ is a discrete probability function not a density. Trivially the conditional expectation

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

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

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

Now we take expectation and get Q function

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

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,

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

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

$$\hat{p}_{j+1} = \frac{1}{n} \sum_{i=1}^n \frac{\hat{p}_j f_1(y_i)}{\hat{p}_j f_1(y_i) + (1-\hat{p}_j) f_2(y_i)} \quad (2)$$

with arbitrary initial estimate $\hat{p}_0 = 0.5$. It will converge quickly after 10 iterations.

Problem 47. Derive iterative scheme similar to (2) for estimating 3 parameters p, μ_1, μ_2 in the two component Gaussian mixture model simultaneously and apply it to the sagittal image. Assume σ_1^2, σ_2^2 are known and equal to the sample variances.

Homework 6. This is the final homework. Solve problems 41- and any other previous unsolved problems. Due: April 30 Friday 11:00am.