Artificial Intelligence	**Homework 7**
CS430 Fall 2003	**Due Wednesday November 19**

1 EM for Gaussians (50 points)

The file `hw7.data` available from the web page contains 40 numbers corresponding to 40 points on the real line. Your task in this assignment is to implement the EM algorithm for fitting a mixture of two univariate Gaussian distributions to the data points.

Denote the 40 numbers as $x_1, \ldots, x_{40}$. Let C be a (hidden) class variable that takes two values 0 and 1. The Bayesian network has the form $C \rightarrow X$. The joint distribution has the form

$$P(X, C) = P(C) \cdot \frac{1}{\sqrt{2\pi}\sigma_C} \exp\left(-\frac{1}{2}\left[\frac{X - \mu_C}{\sigma_C}\right]^2\right)$$

Our goal is to estimate the six parameters $P(C = 0), P(C = 1), \mu_0, \sigma_0, \mu_1$, and σ_1 given only the x's.

We can initialize the algorithm by setting $P(C = 0) = P(C = 1) = 0.5$, $\sigma_0 = \sigma_1 = 1$, $\mu_0 = -1$, and $\mu_1 = 1$.

The **E**-step requires computing $p_{ci} = P(C = c|X = x_i)$ for each example i and class c. From Bayes rule, we know this is

$$p_{ci} = P(C = c|X = x_i) = \alpha_i P(X = x_i|C = c)P(C = c),$$

for some normalizing constant α_i. Hence, let u_{ci} be the un-normalized probability:

$$u_{ci} = \frac{1}{\sqrt{2\pi}\sigma_c} \exp\left(-\frac{1}{2}\left[\frac{x_i - \mu_c}{\sigma_c}\right]^2\right) P(C = c)$$

Then

$$p_{ci} = \frac{u_{ci}}{u_{0i} + u_{1i}}$$

The **M**-step requires re-estimating the 6 parameters of the model. This is done as follows:

$$\begin{aligned} p_c &= \sum_i p_{ci} \qquad \text{Total probability of class } c \\ P(C = c) &= \frac{p_c}{p_0 + p_1} \\ \mu_c &= \frac{1}{p_c} \sum_i p_{ci} x_i \\ \sigma_c &= \sqrt{\frac{1}{p_c} \sum_i p_{ci}(x_i - \mu_c)^2} \end{aligned}$$

Note that in the last equation, you should use the μ_c's computed in the next-to-last equation.

Your program should perform 50 iterations of EM. After each iteration, it should display the six model parameters.

You should turn in a source code listing, a log file showing the execution of the program, and a graph plotting the two means μ_0 and μ_1 as a function of the number of iterations.