ECE 643: Introduction to Biocomputing

Project # 5: On Probability Density Estimation and the EM Algorithm

(Issued Thursday 04/10/2009 – Due Thursday 04/17/2009)

1. Given three coins with the following probabilities of heads
   \[ p_1(h) = \alpha, \quad p_2(h) = \beta, \quad \text{and} \quad p_3(h) = \gamma, \]
   The following experiment is done many times; coin 1 is tossed first, then: If coin 1 turns up heads, coin 2 is tossed twice, and else coin 3 is tossed twice. Estimate the values of \( \alpha, \beta, \) and \( \gamma \) given the following data:
   a. (h,h), (t,t), (h,h), (h,h), (t,t), (t,t)
   b. (h,h), (t,t), (h,h), (h,h), (t,t), (t,t)

2. Consider the \( K \)-component Gaussian mixture model with parameters (K=3)
   \[ \omega_1 = 0.5, \quad \mu_1 = 43, \quad \sigma_1 = 10, \]
   \[ \omega_2 = 0.2, \quad \mu_2 = 128, \quad \sigma_2 = 10, \]
   \[ \omega_3 = 0.3, \quad \mu_3 = 170, \quad \sigma_3 = 10, \]
   a. Write a function that takes as input a value for \( n \) and returns a \( 1 \times n \) vector \( x \) representing a random sample drawn from this mixture model.
   b. Write a function “EM” that takes \( x \) and an integer \( K \) and returns the MLE of a Gaussian mixture model with \( K \) components, as computed by the EM algorithm.
   c. In the code for the EM algorithm, include commands that show the \( K \) different Gaussians components as they are iteratively updated. (Similar to what has been done in the class)
   d. Generate a sample from the Gaussian mixture model of size \( n = 1000 \). Call EM function with \( K = 3 \). Report the following:
      I. The estimates of the model parameters and how they compare to the true values
      II. The effect of initial values and the number of iterations required until convergence
      III. Plot the log-likelihood as a function of iteration number and verify that it is non-decreasing.

3. Using the EM algorithm, estimate the marginal densities of the classes in the following images
   Get an initial segmentation based on these estimations.

4. **Extra credit:** Using the EM algorithm, estimate the marginal densities of the classes in the “rose.ppm” image. Get an initial segmentation based on these estimations.