Homework # 4

1. In this problem, we will revisit the Hope heights problem of HW 3, but this time we will use EM. Recall, we consider the two component Gaussian mixture model,

$$X = \begin{cases} \mathcal{N}(\mu_1, \sigma_1^2) & \text{with probability } p_1 \\ \mathcal{N}(\mu_2, \sigma_2^2) & \text{with probability } p_2 \end{cases} \tag{1}$$

where $\mathcal{N}(\mu, \sigma^2)$ is the normal distribution and X models the height of a person when gender is unknown. Let $\theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, p_1, p_2)$. Let $\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_N$ be the sample heights given in the file.

(a) Recall that to implement EM we define

$$Q(\theta', \theta) = \sum_{i=1}^{N} E_\theta[\log P(\hat{X}_i, z_i \mid \theta')], \tag{2}$$

where z_i is either 0 or 1 and determines the mixture $\hat{X}_i$ was sampled from and the θ subscript in the expectation means that we take the expectation with the z_i distributed according to θ. Write down an expression for $Q(\theta, \theta')$ using $r_{i1} = P(z_i = 1 \mid \hat{X}_i, \theta)$, $r_{i2} = P(z_i = 2 \mid \hat{X}_i, \theta)$ and the pdfs of the normals. Then give a formula for r_{i1} and r_{i2} in terms of θ and the $\hat{X}_i$.

(b) Compute $\text{argmax}_{\theta'} Q(\theta', \theta)$. You should derive an expression for each entry of θ' by solve $\nabla_{\theta'} Q(\theta', \theta) = 0$. (We did μ_1' in class.). Hint: To compute that values of p_1' and p_2' for θ', you can either use a Lagrange multiplier approach, with the constraint $p_1' + p_2' = 1$ or you can simply substitute $p_2' = 1 - p_1'$.

(c) Write an R or Python function to implement the EM algorithm and use it to compute the MLE for the mixture model. To check the correctness of your iteration, show that the log-likelihood increases with each EM iteration. (You can reuse your log-likelihood function from hw 3.)

(d) Given your MLE in (c), use the distribution of X to predict whether a given sample is taken from a man or woman. Determine what percentage of individuals are classified correctly.

2. Let X represent 10 bits, i.e. $X = (X_1, X_2, \dots, X_{10})$ where each coordinate of X is either 0 or 1. Assume the following Bernoulli mixture model for the *ith* coordinate of X, X_i:

$$X_i = \begin{cases} \text{Bernoulli}(\mu_i^{(1)}) & \text{with probability } p_1 \\ \text{Bernoulli}(\mu_i^{(2)}) & \text{with probability } p_2, \end{cases} \tag{3}$$

where $\mu^{(1)}, \mu^{(2)} \in \mathbb{R}^{10}$ with all coordinates in $[0, 1]$. Assume further that the coordinates of X are always sampled from the same mixture, with probabilities p_1 and p_2 for mixture 1 and 2 respectively, but that the Bernoulli draw of each coordinate is independent.

   (a) Write down the EM iteration for this mixture model.

   (b) Attached is the file `noisy_bits.csv` which contains a 500× 10 matrix. Each row of the matrix is a sample of X. If you look at an image of the matrix (in R use **image** on the transposed matrix), you will see that there are two patterns, but with some noise added. Use your EM algorithm to fit the mixture model to the data. Does your fit recover the two underlying patterns?