1. (**"Hard EM" for the k-means clustering algorithm**) Say we have observed data $\mathbf{x} = x_1, x_2, \cdots, x_N$, and we want to cluster this data into K clusters, with means $\mu_1, \cdots, \mu_K$. Let $\mathbf{z} = z_1, z_2, \cdots, z_N$ be our latent variables, which represent the cluster each data point belongs to. Each z_n is a vector, with $z_n^i = 1$ if x_n belongs to cluster i, and 0 otherwise.

   (a) ("E-step") Given an initial guess for the cluster means $\Theta = (\mu_1, \cdots, \mu_K)$, how can the cluster assignments z_n^i be updated?

   (b) ("M-step") Given these new cluster assignments, how can the means be updated?

2. (**EM for a Gaussian mixture model**) Now our model for data generation will be a mixture of K Gaussians. To generate a data point, the ith Gaussian is chosen with probability π_i, then a data point is chosen according to the normal distribution $\mathcal{N}(\mu_i, \sigma_i^2)$. (Here we consider the one dimensional case, but this approach can easily be extended.)

   (a) What is the likelihood of our data?

   (b) (E-step) Let τ_n^i be the posterior probability that x_n came from Gaussian i. Find a formula for τ_n^i.

(c) (M-step) Without taking derivatives, what should the updated parameters μ_i^*, σ_i^*, and π_i^* be?

(d) How does this fit into the equation for EM discussed during lecture?