CS534 — Homework Assignment 6
This is a bonus assignment. If you have not done 80% of the your previous written homeworks, you can use this one to make up.

Due: Friday June 6th.

1. **HAC.** Create by hand the clustering dendrogram for the following samples of ten points in one dimension.

   \[ \text{Sample} = (-2.2, -2.0, -0.3, 0.1, 0.2, 0.4, 1.6, 1.7, 1.9, 2.0) \]

   a. Using single link.
   b. Using complete link

2. **Picking \( k \) for Kmeans.** One shortcoming of Kmeans is that one has to specify the value of \( k \). Consider the following strategy for pick \( k \) automatically: try all possible values of \( k \) and choose \( k \) that minimizes \( J_e \). Argue (briefly) why this strategy is a good/bad idea. Provide an alternative strategy.

3. **Gaussian Mixture Models.** Let our data be generated from a mixture of two univariate gaussian distributions, where \( f(x|\theta_1) \) is a Gaussian with mean \( \mu_1 = 0 \) and \( \sigma^2 = 1 \), and \( f(x|\theta_2) \) is a Gaussian with mean \( \mu_2 = 0 \) and \( \sigma^2 = 0.5 \). The only unknown parameter is the mixing parameter \( \alpha \) (which specifies the prior probability of \( \theta_1 \)). Now we observe a single sample \( x_1 \), please write out the likelihood function of \( x_1 \) as a function of \( \alpha \), and determine the maximum likelihood estimation of \( \alpha \).

4. **Expectation Maximization for Mixture of Multinomials**

   Consider a random variable \( x \) that is categorical with \( M \) possible values \( 1, \ldots, M \). We now represent \( x \) as a vector such that for \( i = 1, \ldots, M \), \( x(i) = 1 \) iff \( x \) takes the \( i \)th value, and \( \sum_{i}^M x(i) = 1 \). The distribution of \( x \) is described by a mixture of \( K \) discrete Multinomial distributions such that:

   \[
   p(x) = \sum_{k=1}^{K} \pi_k p(x|\mu_k)
   \]

   and

   \[
   p(x|\mu_k) = \prod_{j=1}^{M} \mu_k(j)^{x(j)}
   \]

   where \( \pi_k \) denotes the mixing coefficient for the \( k \)th component (aka the prior probability that the hidden variable \( z = k \)), and \( \mu_k \) specifies the parameters of the \( k \)th component. Specifically, \( \mu_k(j) \) represents the probabilities \( p(x(j) = 1|z = k) \), and \( \sum_j \mu_k(j) = 1 \). Given an observed data set \( \{x_i\}, i = 1, \ldots, N \), please derive the E and M step equations of the EM algorithm for optimizing the mixing coefficient and the component parameters \( \mu_k(j) \) for this distribution. For your reference, here is the generic formula for the E and M steps. Note that \( \theta \) is used to denote all the parameters of the mixture model.

   - **E-step.** For each \( i \), calculate \( Q_i(z_i) = p(z_i|x_i;\theta) \), i.e., the prob. that observation \( i \) belongs to each of the \( K \) cluster.

   - **M-step.** Set

     \[
     \theta := \arg\max_{\theta} \sum_{i=1}^{N} \sum_{z_i} Q_i(z_i) \log \frac{p(x_i, z_i; \theta)}{Q_i(z_i)} .
     \]
5. Dimension reduction.

a. Consider the following data set, please draw on the picture the the 1st Principal component direction, and the direction for LDA respectively. Note for PCA, please ignore the markers, and for LDA, we treat the circles as one class and the rest as the other class.

b. Given three data points, (0,0), (1,2), (-1, -2) in a 2-d space. What is the first principal component direction (please write down the actual vector)? If you use this vector to project the data points, what are their new coordinates in the new 1-d space? What is the variance of the projected data?