CS534 — Homework Assignment 4 Due: Saturday Dec 1st. 11:59PM

Please submit your solution in a single pdf file via TEACH.

1. Prove that the kmeans objective
$$J = \sum_{c=1}^{k} \sum_{x_i \in C_c} |x_i - \mu_c|^2$$
monotonically decreases with each iteration of the Kmeans algorithm.

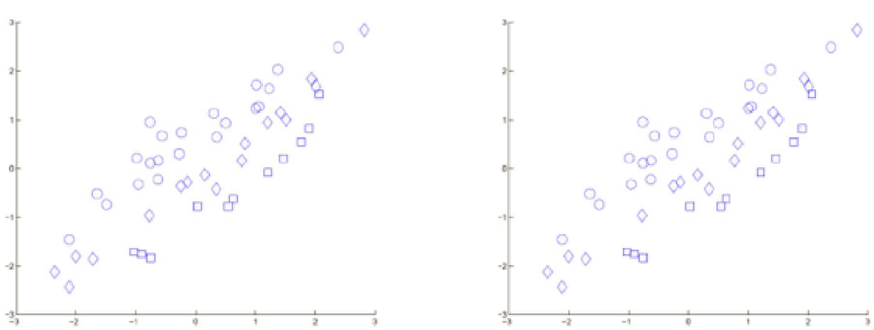
2. **Picking k for Kmeans with J?**. Prove that the minimum of the kmeans objective J is a decreasing function of k (the number of clusters) for $k = 1, ..., n$, where n is the number of points in the dataset. Argue that it is a bad idea to choose the number of clusters by minimizing J.

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 α (which specifies the prior probability of θ_1.). Now we observe a single sample x_1, please write out the likelihood function of x_1 as a function of α, and determine the maximum likelihood estimation of α.

4. **Expectation Maximization for Mixture of Multinomials**
Consider a categorical random variable x with M possible values $1, \cdots, M$. We now represent x as a vector $\mathbf{x}$ such that for $j = 1, ..., M$, $\mathbf{x}(j) = 1$ iff $x = j$. The distribution of $\mathbf{x}$ is described by a mixture of K discrete Multinomial distributions such that:
$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k p(\mathbf{x}|\mu_k)$$
and
$$p(\mathbf{x}|\mu_k) = \prod_{j=1}^{M} \mu_k(j)^{\mathbf{x}(j)}$$
where π_k denotes the prior probability of cluster k, and μ_k specifies the parameters of the kth component. Specifically, $\mu_k(j)$ represents the probabilities $p(\mathbf{x}(j) = 1|z = k)$, and satisfies that $\sum_j \mu_k(j) = 1$. Given an observed data set $\{\mathbf{x}_i\}, i = 1, \cdots, N$, derive the E step and M step for the EM algorithm.

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?