AI534 — Written Homework 4

This assignment covers ensemble methods and clustering.

1. **Boosting.(8 pts)** Please show that in iteration l of Adaboost, the weighted error of h_l on the updated weights D_{l+1} is exactly 50%. In other words, $\sum_{i=1}^{N} D_{l+1}(i) I(h_l(X_i) \neq y_i) = 50\%$, where $I(\cdot)$ is the indicator function that takes value 1 if the argument is true. (Hint: given that the weighted error of h_l is ϵ_l, after the update what is the total weights of incorrectly classified examples? What is the total weights of the correctly classified examples?)

2. **HAC (8pts)**. Create by hand the clustering dendrogram for the following samples of ten points in one dimension.

$$Sample = (-2.2, -2.0, -0.3, 0.1, 0.2, 0.4, 1.6, 1.7, 1.9, 2.0)$$

   a. (4 pts) Using single link.
   b. (4 pts) Using complete link

3. **Kmeans with L_1 norm (10 pts)**. Consider replacing the distance function used for Kmeans with L_1 norm, which gives us the following objective:

$$\min_{\mu_1,\ldots,\mu_K, C_1,\ldots,C_K} \sum_{i=1}^{K} \sum_{\mathbf{x} \in C_i} |\mathbf{x} - \mu_i|$$

   (a) (5 pts) Show that given fixed cluster assignments $C_1, \ldots, C_K$, the prototype μ_i that optimizes the above objective can be obtained by taking the median of each dimension for cluster i (Hint: use the fact that the derivative of the function $f(x) = |a - x|$ is 1 if $x > a$ and -1 if $x < a$.)

   (b) (3 pts) Modify the kmeans algorithm for this L_1 based objective.

   (c) (2 pts) Comparing this algorithm with the regular K-means algorithm, which one is more robust to outliers? Why?

4. **Picking k for Kmeans with J? (6 pts)**. Prove that the minimum of the kmeans objective J is a decreasing function of k (the number of clusters) for $k = 1, \ldots, n$, where n is the number of points in the dataset. Explain why it is a bad idea to choose the number of clusters by minimizing J.

5. **Gaussian Mixture Models in 1-d (8 pts)**. Let our data be generated from a mixture of two 1-d 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.

6. **Expectation Maximization for Mixture of Categorical distributions (bonus 10 pts)** 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 $j = 1, \ldots, M$, $\mathbf{x}(j) = 1$ iff $x = j$. The distribution of $\mathbf{x}$ is described by a mixture of K discrete categorical 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 distribution of the k-th cluster. 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 $\{\mathbf{x}_i\}, i = 1, \cdots, N$, Please write out the E step and M step for the EM algorithm for learning the mixture of categorical distributions.