1. **Boosting.** Consider a data set \( D = x_1, x_2, ..., x_{10} \). We apply Adaboost with decision stump. Let \( w_1, \ldots, w_{10} \) be the weights of the ten training examples respectively. In the first iteration, \( x_1, x_2 \) and \( x_3 \) are misclassified. Please rank the updated weights \( w_1, w_2, ..., w_{10} \) in increasing order. In the second iteration, \( x_3 \) and \( x_4 \) are misclassified. Please provide the rank of the updated weights \( w_1, w_2, ..., w_{10} \) in increasing order. Explain your ordering.

2. **MAP estimation.** Consider the problem of linear regression. We are given a set of observed data points \((X^i, t^i), i = 1, ..., N\), where \( X \) is the input vector, and \( t \) is the target output. The goal is to estimate a set of linear coefficients \( W \) such that \( t \) can be predicted by \( W^T X \). In particular, we assume that \( t \sim N(W^T X, \sigma^2) \). Assume that each coefficient \( w_i \) has a prior distribution \( N(0, \alpha^{-1}) \). Please write down the posterior likelihood function of \( W \), and show that maximizing this posterior likelihood is equivalent to minimizing the least square objective (3.12) with a L2 regularization term.

3. **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 why this strategy is a good/bad idea. Provide an alternative strategy.

4. **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 \( \alpha \).

5. **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

Note that in class we described HAC assuming similarity functions. You should be able to easily revise the definitions to use distance functions instead. And here the Euclidean distance function is used.