1. **HAC.** Create a 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 please use the Euclidean distances.

**Solution:**

*For the single link case we get the following:*

![Single Link Dendrogram]

*For the complete link case, the hierarchy remains the same but the height of the dendrogram is different.*

![Complete Link Dendrogram]
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 why this strategy is a good/bad idea.

**Solution:** This strategy is a very bad idea. The reason is that the optimal distortion on the training data will always decrease as we increase \( k \) (until \( k \) equals the number of instances). In particular, note that when \( k = n \) the number of instances that \( J_e = 0 \) so the approach will always select \( k = n \).

3. **Gaussian Mixture Models.** Let \( f(x) \) be a univariate mixture density such that \( f(x|\theta_1) \) is a Normal distribution with mean \( \mu_1 = 0 \) and \( \sigma^2 = 1 \), and \( f(x|\theta_2) \) is a Normal distribution with mean \( \mu_2 = 0 \) and \( \sigma^2 = 0.5 \). Thus we can write \( f(x) \) as:

\[
f(x) = \frac{\alpha}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2} + \frac{1-\alpha}{\sqrt{\pi}}e^{-x^2}
\]

where \( \alpha \) is the unknown parameter specifying the prior probability of \( \theta_1 \). Consider that a single sample \( x_1 \) has been observed. Determine the maximum likelihood estimate of \( \alpha \).

* I will present two different approaches to solve this problem:
**EM approach**

Initial guess. $\alpha = 0.5$

**Iteration 1**

**E-step:** $P(Y=1|X_i) = \frac{1}{\sqrt{2\pi}e^{\frac{x_i^2}{2}}} \cdot \frac{e^{-\frac{x_i^2}{2}}}{a + b} = \frac{a}{a + b}$

$P(Y=2|X_i) = \frac{b}{a + b}$

**M-step:** $\alpha = \frac{a}{a + b}$

**Iteration 2**

**E-step:** $P(Y=1|X_i) = \frac{1}{\sqrt{2\pi}e^{\frac{x_i^2}{2}}} \cdot \frac{a}{atb} \cdot \frac{e^{-\frac{x_i^2}{2}}}{atb} = \frac{a^2}{a^2 + b^2}$

$P(Y=2|X_i) = \frac{b^2}{a^2 + b^2}$

**M-step:** $\alpha = \frac{a^2}{a^2 + b^2}$

It can be shown that, at iteration $t$, $\alpha = \frac{a^t}{a^t + b^t}$.

If $a < b$, as $t \uparrow$, $\alpha$ converges to zero.

If $a > b$, as $t \uparrow$, $\alpha$ converges to one.

If $a = b$, $\alpha = \frac{1}{2}$ and EM converges at step 2.