**1. Theory Questions**

1.

Cluster 1

Cluster 2

**a. The weighted average intra-cluster distance if you are using Euclidean distance?**

Cluster 1:

Cluster 2:

Weighted Average Intra-cluster distance

**b. The single link similarity between the clusters if we're using cosine similarity as our similarity function?**

Single link means comparing the most similar: (1,2) and (1,1)

**c. The complete link similarity between the clusters if we're using cosine similarity as our**

**similarity function?**

Complete link means comparing the furthest apart: (1,1) and (0,-1)

**d. The average link similarity between the clusters if we're using cosine similarity as our**

**similarity function?**

Average Link means comparing the average of the two points in the cluster

**2. Given an average intracluster distance for clustering level j, Wj, what is the fourth derivative at j, namely Wj’’’’**

We can break the Wj’’ given in the notes into parts, and use the same pattern to get the second derivative of those parts.

**3. What is the weighted average purity of the clusters created by the clustering algorithm?**

First, calculate all the possible Nij’s

N11 = 2, N12 =2

N21 = 2, N22 = 4

Now Purity of Each Cluster

Now Average Purity