

**CSCI 6350/7350 Spring 2012****Homework 5 (due beginning of class, Monday, March 12<sup>th</sup>)**

1. Calculate the dissimilarity between pair-wise objects using the dissimilarity measure discussed in class. List the dissimilarities in 6x6 dissimilarity table. Fever and Cough are asymmetric binary attributes, Category is ordinal attributes with the order of the three values being: stable, unstable and severe.

	<b>Gender</b>	<b>Age</b>	<b>Heart-rate</b>	<b>Fever</b>	<b>Cough</b>	<b>Category</b>
Obj1:	M	18	120	Y	N	stable
Obj2:	F	36	89	N	Y	unstable
Obj3:	M	20	115	Y	Y	stable
Obj4:	M	3	94	Y	N	severe
Obj5:	F	28	110	N	Y	severe
Obj6:	F	44	80	N	Y	unstable

2. Applying PAM clustering approach on the above data to partition data into K=2 clusters. Assuming Obj2 and Obj4 are selected as the initial medoids of the two clusters. Show:
  - a. Which cluster would each of the data objects be assigned to on the first iteration of object distribution?
  - b. Should Obj3 replace Obj4 as the cluster medoid in the 2nd iteration of clustering? Answer the question with results from computation.
3. Perform hierarchical clustering on the six objects in question 1. Use the average link agglomerative clustering methods. Show the clustering hierarchy constructed.
4. Apply Kmeans clustering algorithm to group the following 6 objects into K=3 clusters. (Show all steps in all iterations before the clusters converge. Assuming obj1, obj3, and obj5 were selected initially as the cluster centroids for the three clusters. (no need to do normalization in this problem)

	<b>length</b>	<b>width</b>	<b>height</b>
obj1:	3	18	5
obj2:	5	12	10
obj3:	1	16	7
obj4:	4	12	9
obj5:	8	5	8
obj6:	9	3	6

5. Read the paper : BIRCH: An Efficient Data Clustering Method for Very Large Databases. (downloadable from the course web site).