

**Homework 8 Due beginning of class, Tuesday, April 15<sup>th</sup>**

1. Calculate the dissimilarity between pair-wise objects using the dissimilarity calculation formula developed for variables of mixed type (refer to slide 31 in clustering (1) file). 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.

Note: This computes the **dissimilarity values** between pairwise objects

	<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. In which cluster would each of the data objects be assigned on the first iteration of object distribution?
  - b. Should Obj 3 be used to replace Obj4 as the Medoid of one of the clusters for the next iteration of clustering? Answer the question with results obtained from computation.
3. Perform hierarchical clustering on the six objects in question 1, using the **average-link** agglomerative clustering methods. Show the intermediate similarity clustering hierarchy constructed.
4. Given the following transaction database, apply **Apriori** algorithm to derive from the database:
  - (a) all frequent itemsets, and
  - (b) all strong association rules
 that has the min\_support\_count = 2, and the min\_confidence 70%.  
 Show all intermediate steps ( $C_k$ ,  $L_k$ , Join steps, ...) as done in class.

TID	Items bought
T100	{K, A, D, B}
T200	{D, A, C, E, B}
T300	{C, A, B, E}
T400	{B, A, D}