Assignment 3 (due to Wednesday, June 14)

1. Giving the distance matrix of n objects $D=(d_{ij})_{n \times n}$, where $d_{ij} = d_{ji}$ stands for the distance between objects i and j.

$$S = \begin{bmatrix} 0 & 1.3 & 1.0 & 1.0 & 1.3 & 1.3 & 0.7 \\ & 0 & 1.2 & 1.2 & 0.1 & 0.4 & 1.3 \\ & & 0 & 0.5 & 1.3 & 1.2 & 1.0 \\ & & & 0 & 1.3 & 1.3 & 1.0 \\ & & & & 0 & 0.2 & 1.2 \\ & & & & & 0 & 1.2 \\ & & & & & & 0 \end{bmatrix}$$

a) Perform the agglomerative (bottom-up) hierarchical clustering with single link (M.Sc) or complete link (Ph.D)

b) Compute the cophenetic matrix C of your clustering in (a) and evaluate your clustering in a in terms of r(C,S)

2. Assume that 6000 genes are clustered by two methods, one results in 5 clusters, another results in 6 clusters. The following is the contingency table

	v1	v2	v3	v4	v5	v6	total
u1	25	50	45	900	30	0	1050
u2	75	1000	40	20	45	20	1200
u3	20	25	550	30	25	700	1350
u4	900	20	10	20	40	10	1000
u5	50	25	25	30	1250	20	1400
total	1070	1120	670	1000	1390	750	6000

a) Calculate the ARI

b) Calculate MI