CAP5510 Fall 2019, HW#3, Due Wednesday, Nov 6th, 2019 (hard copy submission in class)

Total 30 pts

1. (15 pts) Given the following expression matrix in which rows represent genes and columns represent experiment conditions,
(1) Compute the correlation matrix for the 5 genes based on Pearson's correlation estimation;
(2) Cluster genes using hierarchical clustering based on Euclidean distance and Centroid linkage;
(3) Show the first 2 rounds of k-means clustering based on Euclidean distance, where k=2 and set the initial vectors to be the expression vectors of p53 and bcl2.

Exp.Values	Condition1	Condition2	Condition3	Condition4	Condition5
P53	10	4	2	8	6
Mdm2	9	1	1	7	5
Bcl2	2	10	4	5	9
cyclinE	7	6	5	6	6
Caspase 8	2	10	6	4	8

2. (15 pts)
We have learned in class two clustering algorithms such as hierarchical clustering and k means clustering. As we know, both algorithms are not perfect. Please identify one new clustering algorithm that was published after 2016:
(1) explain how it works and use one toy example to illustrate the steps.

(2) explain its advantages and disadvantages comparing with hierarchical and k-means algorithms using one toy example.