

CS 481, First Midterm Exam (Spring 2016)

Time: 75 minutes, Total: 75

Name: _____

Questions

1: Given the following points: 2, 4, 6, 9, 15, 18. Assume $k = 2$, and that we randomly pick the initial means $\mu_1 = 2$, and $\mu_2 = 6$. Now answer the following questions.

- Show the cluster obtained using k-means algorithm after 2 iterations. You need to show cluster membership and the new mean for each iteration. (15)
- Now consider EM clustering. For the same means, if we assume $\sigma_1=2$, and $\sigma_2=1$, $P(C_1) = 1/3$ and $P(C_2) = 2/3$, Compute $P(C_1|x = 4)$, and $P(x = 6|C_2)$. (10)

2: For the following clustering algorithms, (1) k-Means, (2) Graph Clustering, (3) EM Clustering write down their objective functions along with a short justification (15)

3: For each of the following clustering methods, list their pros and cons (at least 2 properties for each category): (1) k-Means, (2) hierarchical clustering, (3) graph clustering (10)

4: A clustering method yields 3 clusters, C_1 , C_2 , and C_3 from a dataset. The dataset contains 3 categories of objects, namely T_1 , T_2 , and T_3 . Given the following contingency table, compute F-measure of each clusters. (10)

	T_1	T_2	T_3
C_1	5	75	10
C_3	15	5	50
C_2	55	10	5

5. Consider the graph below where each edge is labeled with a similarity score. For the following two clustering $\mathcal{C}_1 = \{(A, B, C), (D, E, F, G)\}$ $\mathcal{C}_2 = \{(A, B, C, D), (E, F, G)\}$ which is better; argue by computing the the ratio-cut and normalized cut score. (15)

