Submitted by: Sheikh Mastura Farzana, Md. Emtazul Haque
Cüneyt Erem

2. Convergence of k-means

a) If at timestep t, k-means is not at final state, two iterationsteps can change μ between t and t+1:

step 3: At least one point will be reassigned to a new and closer cluster center. Hence, the sum of distances between points and their cluster centers will decrease, too.
$\rightarrow \mu_t > \mu'_{t+1}$ ✓

0.5/1

step 4: Since all points have the same weight, repositioning the cluster centers at the center of weight of their cluster does not increase the sum of distance. squared
$\rightarrow \mu'_{t+1} \geq \mu_{t+1} \Rightarrow \mu_t > \mu_{t+1}$

You need to prove this!

1/1 b) For n points and k clusters, there are k possible assignments for each point, resulting in a maximum of k^n possible assignments. OK.

0/1 c) The two phases of re-assigning data points to clusters and re-computing the cluster means are repeated in turn until there is no further change in assignments or until some maximum number of iterations $\leq k^n$ is exceeded. Hence finiteness is guaranteed.

no. it is repeated until the assignment doesnt change and this always happens after $< k^n$ steps as the same assignment results in the same error and the error only decreases so each assignment can only occur once.

8. Expectation Maximization

a) The runtime of the EM algorithm is changed by any modification to the parameters of the initial parameters. The algorithm will take less time to run if initial cluster parameters are close to a local minimum. EM makes assignment based on the posterior probabilities. It is a two-stage iterative optimization technique for finding maximum likelihood solutions. These solutions are bounded by 0 which implies local optimization. OK

1/1

b) For a big value of k, they might cover a big space and may not contain points. For small k clusters, they might be combined to form big clusters and the algorithm might loose granularity. over-/underfitting

1/1

c) For better convergence the value of the likelihood function needs to be increased. The initial parameters of the EM algorithm can be preprocessed and pre computed for optimal result. Different hyper parameter optimization techniques can be useful for this. ✓

1/1

d)

0/1

Theory Points

0.5+1+0+1+1+1+0 / 7

Practical Points

3+1+0.5 / 5

c- weird result (clusters inside each other)