HW10 M5

Answer (for reference only)

1.

Prove that K-Means algorithm always converges in finite number of steps.

K-Means algorithm works as follows:

1. Initialize
   - Select initial *m* cluster centers
2. Find associations
   - For each x_i, assign the cluster with nearest center
   - ➔ Find A to minimize J(X; C, A) with fixed C
3. Find centers
   - Compute each cluster center as the mean of data in the cluster
   - ➔ Find C to minimize J(X; C, A) with fixed A
4. Stopping criterion
   - Stop if clusters stay the same. Otherwise go to step 2.

Proof.

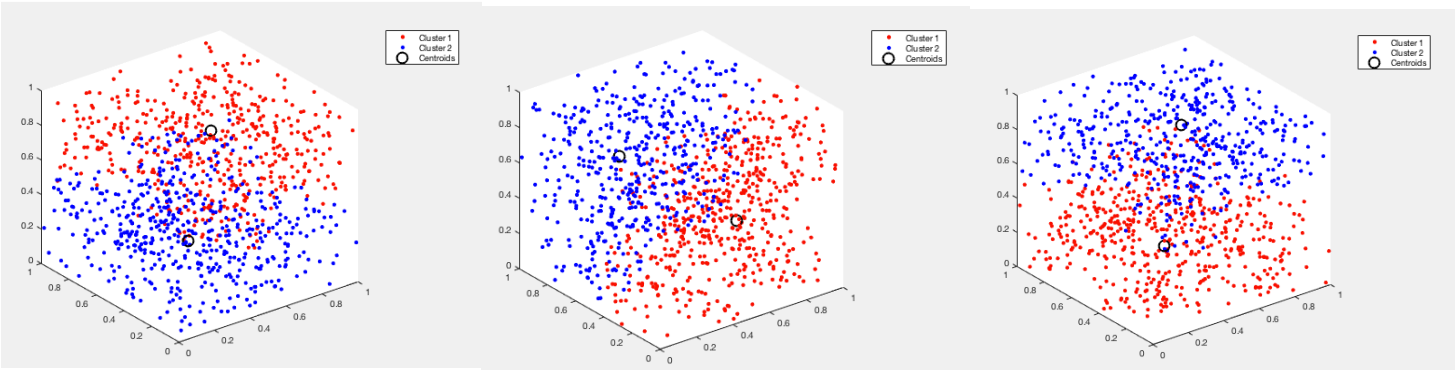
First, we want to choose n centroids from k clusters, that is ay most k^n solutions, which is finite. Then, according to step 2 and 3, we can get $J(X; C, A) \geq J(X; C', A')$, which means that after every iteration, the result will be smaller or equal to the previous result. After summarizing the points above, we can get two different situations:

(1) $J(X; C, A) = J(X; C', A')$ ➔ Terminate

(2) $J(X; C, A) > J(X; C', A')$ ➔ Execute at most k^n times.

In conclusion, K-Means always converges in finite number of steps.

2.



Because x, y, and z is uniformly distributed between 0 and 1, no matter what the input data is, K-Means would just simply divide your data into K clusters. There are no actual clustering phenomena, which means that what K-Means outputs are meaningless.