***Clustering Quiz***

* For which of the following tasks might K-means clustering be a suitable algorithm? Select all that apply.
* **Given a database of info about users, automatically group them into different market segments**
* **Given sales data from a large number of products in a supermarket, figure out which products tend to form coherent groups (say are frequently purchased together) and thus should be put on the same shelf.**
* Suppose we have three cluster centroids:



Furthermore, we have a training example



After a cluster assignment step, what will c(i) be?

* **c(i)=3** (minimized norm of x(i) – u 🡪||x(i) – u||^2)
* K-means is an iterative algorithm, + 2 of the following steps are repeatedly carried out in its inner-loop. Which 2?
* **Move the cluster centroids, where the centroids μk are updated.**
* **The cluster assignment step, where the parameters c(i) are updated.**
* Suppose you have an unlabeled dataset {x(1),…,x(m)}. You run K-means with 50 different random initializations, and obtain 50 different clustering sets of the data. What is the recommended way for choosing which one of these 50 clustering set to use?
* **Compute the distortion function J(c(1),…,c(m),μ1,…,μk), and pick the one that minimizes this.**
* Which of the following statements are true? Select all that apply.
* **If worried about K-means getting stuck in bad local optima, 1 way to ameliorate (reduce) this problem is if we try using multiple random initializations.**
* **For some datasets, the "right" or "correct" value of K (the number of clusters) can be ambiguous, and hard even for a human expert looking carefully at the data to decide.**