Questions

1. How do you characterize the termination of the algorithm, if, say, k = 3? The question is deliberately open-ended since overspecifying the question gives the game away but we are referring to this statement. The algorithm has converged when the assignments no longer change.

Cluster centers are reassigned after every iteration in the k-means clustering algorithm. For k =3, if we start with three randomly selected points and create the cluster set for every center. The mean of the cluster set is the corresponding center for the next iteration. Over a series of iterations, when we observe there is no significant shift in the cluster centers of all the three clusters, then we terminate the algorithm. Basically, when the euclidean distance between the previous and the current assignment is lower than some threshold.

2. For the same input data, can two different valid implementations generate different solutions, assuming both converge?

Yes. If the initializations for the two implementations are drastically different, then they can converge to different solutions.