Problem 1: (5 points) In this problem, you will show that the K-means algorithm is guaranteed to converge to a local minimum. Do this by using the following steps:

(a) Let  $\{X_i\}_{i=1}^n$  be observed samples in  $\mathbb{R}^p$ . Recall the K-means objective

$$\mathcal{L}(\mathcal{W}) = \sum_{k=1}^{K} \frac{1}{|W_k|} \sum_{i,j \in W_k} \|X_i - X_j\|^2.$$

 $\mathcal{W} = (W_1, \dots, W_K)$  is a set a clusters, where  $W_k$  is the set of i such that  $X_i$  is in cluster k. Let  $\mathcal{W}^{(t)}$ denote the clusterings after t steps of K-means. Show that

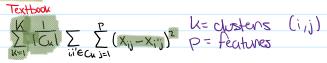
$$\mathcal{L}(\mathcal{W}^{(t+1)}) \leq \mathcal{L}(\mathcal{W}^{(t)})$$

Hint: Rewrite L(W) using the mean vector like we did in class (see Equation 12.18 in Chapter 12 of the ISLRv2 textbook), and show that each step of K-means must decrease



$$L(W) = \sum_{k=1}^{K} 2 \sum_{i,j \in W_k} \|x_i - x_k\|^2$$

After each step in L(W) the k-means algorithm will assign an observation to the one with the least evolidean distance,  $\|\mathbf{x}_i - \mathbf{x}_k\|^2$ . Such that every step decircuses objective.



$$\frac{1}{|C_{k}|} \sum_{i,i' \in C_{k,j=1}}^{P} (x_{ij} - x_{i'j})^{2} = 2 \sum_{i \in C_{k}} \sum_{j=1}^{P} (x_{ij} - \overline{x}_{kj})^{2}$$

Algorithm 122 cleaneuse objective each step

Iterate until the duster assignments stop changing

- (a) For each cluster compute cluster controid.

  The kth cluster centroid is the vector of the p feature means for the observations in the kth cluster
- (b) Assign each observation to the cluster whose Centroid is closest.

(eudidean distance)

12.17

minimize
$$\sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in \mathcal{U}} \sum_{j=1}^{p} \frac{\text{centroid distance}}{(x_{ij} - x_{i'j})^2}$$
We have a substance

(b) Argue that the K-means algorithm will eventually exactly converge to a local minimum. Hint: Use the fact the set of possible K-means outcomes is finite.

We know from equation 12.18 that  $\|x_i - x_i\|^2$  is the teature minus the closter mean for said teature used to minimize sum of savened deviations. We know relocating the algorithms can only improve as shown in (a).

This means that as the algorithm runs, clustening will continuously improve until it can notongen. This is because set of possible K-means outcomes is finite. Meaning it will converge at a local minimum,