PART 1 : K-MEANS

1.2 Question 1

let $c_j = \dfrac{\sum_{i=1}^{\ell} r_{ij} x_i}{\sum_{i=1}^{\ell} r_{ij}}$, $j = 1, \dots, k$

where $r_{ij} = \begin{cases} 1, & \text{if } j = \arg\min_{1 \le s \le k} \|x_i - c_s\|^2 \\ 0, & \text{else} \end{cases}$

and $c_i = \arg\min_{z \in \mathbb{R}^n} \sum_{x \in C_i} \|z - x\|^2$. (1)

$$\frac{\partial}{\partial z} \sum_{z \in \mathbb{R}^n} \|z - x\|^2 = \sum_{x \in C_i} \frac{\partial}{\partial z} \|z - x\|^2 = \sum_{x \in C_i} 2(z - x)$$

set $\sum_{x \in C_i} 2(z - x) = 0 \Rightarrow \sum_{x \in C_i} z = \sum_{x \in C_i} x \Rightarrow \hat{z} = \dfrac{\sum_{x \in C_i} x}{\# C_i}$

<u>Note</u>: We have define $r_{ij} = 0, 1$ as an indicator function describing which cluster $\underline{C_j}$ the data point $\underline{x_i}$ belongs then in the Update step where we recompute the clusters mean to be the new centers of the clusters

$$c_j = \frac{\sum_{i=1}^{\ell} r_{ij} x_i}{\sum_{i=1}^{\ell} r_{ij}.}$$

$\leftarrow$ # of x_n belongs in a specific cluster

$\leftarrow$ # of 1 i.e: x_i belongs c_j.

$$\hat{z} = \frac{\sum_{x \in C_i} x}{\# C_i}$$

$\leftarrow$ we sum over the clusters

Henceforth using the indicator r_{ij} we can write z as c_j. Thus $\hat{z} = c_j$, i.e: the centroid is the minimizer of the sum of square distances (1).

PART 1: K-MEANS

1.2 Question 2

Overview: Suppose we have a data set $x_1, \dots, x_\ell$ our goal is then to give a disjoint partition into k sets $C_1, \dots, C_K$. Assume that we apriori know how many clusters the data should divide. Introduce prototypes $c_1, \dots, c_K$ where each c is a prototype associated with one of the k-Clusters. The prototype represents the center of the cluster.

Goal: To find the $c_1, \dots, c_K$ centers that partition the data in such the sum of the euclidean distances is minimal.

Non-Convex Optimization Problem: $\underset{C_1,\dots,C_K;\, c_1,\dots,c_K}{\arg\min} \sum_{i=1}^{K} \sum_{x \in C_i} \|x - c_i\|^2$

Method: first assign some initial random position to the centers $c_1, \dots, c_K$. Given the data set $x_1, \dots, x_\ell$ and a k-number of clusters, create a set of binary $r_{ij} = 1, 0$ indicator variables describing which cluster c_j the data point x_i belongs to, i.e.: if x_i is assigned to cluster j then $r_{ij} = 1$ otherwise $r_{is} = 0$ for $s \neq j$. Next the k-means algorithm proceeds by alternating between two steps: assignment and update.

Assignment Step: $r_{ij} = \begin{cases} 1 & , \text{if } j = \underset{1 \le s \le k}{\arg\min} \|x_i - c_s\|^2 \\ 0 & , \text{else} \end{cases}$

Update Step: Recompute the cluster means to the new centers of the clusters $c_j = \dfrac{\sum_{i=1}^{\ell} r_{ij} x_i}{\sum_{i=1}^{\ell} r_{ij}}$ $(j = 1, \dots, k)$

- Each iteration (kn) operations

Proof of Corvergence:

- The k-means problem can be represented as the following non convex optimization problem;

$$\underset{C_1,\dots,C_k, c_1,\dots,c_k}{\arg\min} \sum_{i=1}^{k} \sum_{x \in C_i} \| x - c_i \|^2 \quad (1)$$

- We can consider the (1) as a cost function.

$$cost(c_1,\dots,c_k) = \underset{c_1,\dots,c_k}{\arg\min} \sum_{i=1}^{k} \sum_{x \in C_i} \| x - c_i \|^2$$

- From the definition of the k-means algorithm it is clear that each iteraction decreases the cost function. Hence since this is done repeatidly. the loss is decreasing monotonically.

  - Let $C_1^{old},\dots,C_k^{old}$ be the old clusters and. $C_1^{new},\dots,C_k^{new}$ be the new clusters.

  Hence $cost(C_1^{old},\dots,C_k^{old},\underset{\sim}{c}) \geq \underset{C_1^{old},\dots,C_k^{old}}{\min} (C_1^{old},\dots C_k^{old},\underset{\sim}{c}) = cost(C_1^{new},\dots,C_k^{new},\underset{\sim}{c})$

  - Do the same for the centroids let $c_1^{old},\dots,c_k^{old}$ be the old centroids and. $c_1^{new},\dots,c_k^{old}$ be the new centroids.

  Hence $Cost(C_1^{old},\dots,c_k^{old},\underset{\sim}{C}) \geq \underset{c_1^{old},\dots,c_k^{old}}{\min}(C_1^{old},\dots,c_k^{old},\underset{\sim}{C}) = Cost(c_1^{new},\dots,c_k^{new},\underset{\sim}{C})$

- Thus the cost is indeed monotonically decreases. Hence for n iterations (until convergence) the K-means will Convergence, to a local minimum. (<u>Not</u> Global). where $n<\infty$. Finally $n<\infty$ because of the finite number of the updates.