Modern Machine Learning
Homework #6

1. Suppose we have the following dataset

X	9	8	7	7	6	5	5	4	3	2	2	1
Y	8	5	7	2	7	3	4	2	2	4	3	1

and two cluster centroids located at

X	3	8
Y	1	4

  (a) Compute 4 iterations of K-means

  (b) Draw a figure for each iteration, showing the cluster assignment and the positions of the centroids.

2. Prove that the K-means algorithm converges in a finite number of steps.

3. The traditional K-means algorithm is based on the L_2 norm. Provide a K-means variation that utilizes the L_1 norm. Provide advantages/disadvantages of this approach.

   Hint: the definition of L_1 median is

$$\hat{\mu} = \arg\min_{\mu} \sum_i ||x_i - \mu||_1 \tag{1}$$

   where $\hat{\mu}$ denotes the optimal L_1 median centroids.

4. *SoftMax clustering.* In traditional K-means, we employ the L_2 norm, $||\mathbf{x}_n - \boldsymbol{\mu}_k||$, between an observation sample $\mathbf{x}_n$ and a cluster centroid $\boldsymbol{\mu}_k$, where N and K are the number of samples and clusters, respectfully. Suppose we use a soft assignment for each sample relative to a particular cluster centroid. In that case, we can use the responsibility for each sample relative to a particular cluster centroid. Here we use the SoftMax

$$r_{n,k} = \frac{\exp(-\beta||\mathbf{x}_n - \boldsymbol{\mu}_k||)}{\sum_{i=1}^{K} \exp(-\beta||\mathbf{x}_n - \boldsymbol{\mu}_i||)}$$

(a) Determine the expressions and steps necessary to optimize such a SoftMax clustering algorithm.

(b) Determine the role β plays in the algorithm. Determine an appropriate way to set the value of this parameter.

(c) Determine the advantages and disadvantages of this clustering algorithm versus traditional K-means.