Tutorial 6 : K-Means

1 K-Means Algorithm

This is an iterative algorithm which partitions a set of vectors (samples) into K groups, and for each group it finds a 'center of mass' point, or a centroid, which represents the group. We assume for now that the number of groups K is given.

Notation: μ_i is the centroid of the group $G_i,\ i = 1, 2, ..., K$.

The Algorithm:

- Initialization - Choose K centroids $\{\mu_i\}_{i=1}^K$, set $t = 0$.
- Do -
  1. Classify the points using $1 - NN$ algorithm with respect to the centroids. That is, associate a point x to group $G_i^{(t)}$ if

     $$i = \underset{j=1,2,...,K}{\arg\min} \quad ||x - \mu_j^{(t)}||_2$$

     When there is more than one minimum, choose the group with the smallest index.
  2. Compute the new centroids

     $$\mu_i^{(t+1)} = \frac{1}{|G_i^{(t)}|} \sum_{x \in G_i^{(t)}} x$$

     where $|G_i^{(t)}|$ denotes the number of elements in $G_i^{(t)}$. When $|G_i^{(t)}| = 0$ set $\mu_i^{(t+1)} = \mu_i^{(t)}$.
  3. Set $t \leftarrow t + 1$ and go to set (1) until convergence

     $$\mu_i^{(t+1)} \approx \mu_i^{(t)},\ i = 1, 2, ..., K.$$

The algorithm minimizes the sum of squared errors

$$\sum_{i=1}^{K} \sum_{x \in G_i} ||x - \mu_i||_2^2.$$

The algorithm is guaranteed to converge to a local minimum. Empirically it is observed that the algorithm is robust to initialization. However, it is recommended to choose the initial centroids in a thoughtful manner, using prior knowledge if possible.

Setting K

In most cases we do not know K in advance, but want to find some reasonable choice for K. The clustering error for a given K is defined as

$$E(K) = \sqrt{\sum_{i=1}^{K} \sum_{x \in G_i} ||x - \mu_i||_2^2}.$$

As we increase K, the error will become smaller. If we set K to be the number of samples, each sample will get its own cluster, the error will be zero, but we did not learn anything. One possible method to choose a reasonable number of groups, is to increase K gradually and to compute $E(K)$ at the end of each step. We stop increasing K when

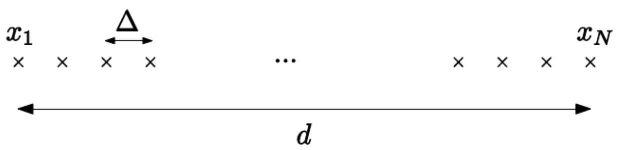
$$1 - \frac{E(k)}{E(K-1)} \leq \epsilon$$

for some threshold ϵ.

2 Practice

Question 1

Consider the following 1-dimensional clustering problem:



where the samples $\{x_i\}_{i=1}^{N}$ are positioned uniformly on the interval $[0, d]$, and their number $N \to \infty$ (and of course $\triangle \to 0$).

Show that the K-Means algorithm with $K = 2$ converges to the global minimum of the squared error, from any **reasonable initial condition**, that is, the initial centroids are located in the interval $[0, d]$.

Solution

Let $\mu_1^{(0)}$ and $\mu_2^{(0)}$ be the initial centroids of group G_1 and G_2 respectively.
We denote by $x^{(0)}$ the boundary point for which

$$x \leq x^{(0)}, \ \forall x \in G_1,$$
$$x > x^{(0)}, \ \forall x \in G_2.$$

Then, $x^{(0)}$ is given by

$$x^{(0)} = \frac{\mu_1^{(0)} + \mu_2^{(0)}}{2} = \alpha d$$

for some $0 \leq \alpha \leq 1$. At the first iteration we get

$$\mu_1^{(1)} = \frac{1}{2} x^{(0)},$$
$$\mu_2^{(1)} = \frac{1}{2}(x^{(0)} + d),$$
$$x^{(1)} = \frac{\mu_1^{(1)} + \mu_2^{(1)}}{2} = \frac{1}{2} x^{(0)} + \frac{d}{4}.$$

In general,

$$\mu_1^{(n)} = \frac{1}{2}x^{(n-1)},$$
$$\mu_2^{(n)} = \frac{x^{(n-1)} + d}{2},$$
$$x^{(n)} = \frac{1}{2}x^{(n-1)} + \frac{d}{4}.$$

The algorithm converges, hence, $x^{(n)} \to x^* \in [0, d]$. We can find x^* using the recursion above

$$x* = \frac{1}{2}x^* + \frac{d}{4} \;\Rightarrow\; x^* = \frac{d}{2}.$$

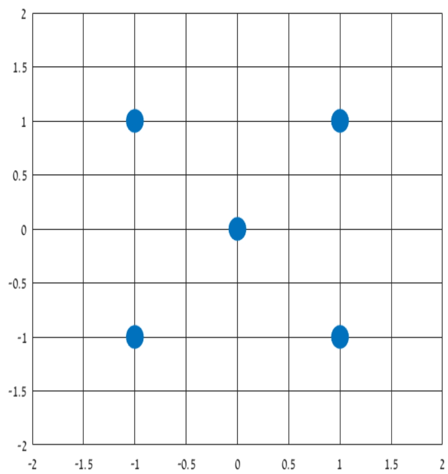
Therefore, the solution is

$$\mu_1 = \frac{d}{4},\; G_1 = \left\{x : x \le \frac{d}{2}\right\},$$
$$\mu_2 = \frac{3d}{4},\; G_2 = \left\{x : x > \frac{d}{2}\right\}.$$

Question 2

Consider a set of five examples:

$$x_1 = \begin{bmatrix}1\\1\end{bmatrix},\; x_2 = \begin{bmatrix}1\\-1\end{bmatrix},\; x_3 = \begin{bmatrix}-1\\-1\end{bmatrix},\; x_4 = \begin{bmatrix}-1\\1\end{bmatrix},\; x_5 = \begin{bmatrix}0\\0\end{bmatrix},$$



Recall that K-means algorithm changes the classification of a sample to another group only when its distance to the centroid of that group is strictly smaller than the centroid of its current group.

(a) For $K = 2$, find all possible partitions which the algorithm can converge to.

(b) For each one of the partitions you found, state whether the partition is robust or not. A partition is considered robust if applying a small change on one of its centroids and running the algorithm again, results in the same partition.

Solution

(a) There are 12 possible partitions:

- One group is empty and the second includes all examples and its centroid is at $(0, 0)$. (1 option).
- One group has only one example and the second includes all other four
  i. The first group includes one of the vertices and the second includes all other four points. For example, $G_1 = \{x_3\}$ with $\mu_1 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$ and G_2 includes all other examples with $\mu_2 = \begin{bmatrix} 1/4 \\ 1/4 \end{bmatrix}$. (4 options by symmetry)
  ii. The first group includes $x_5 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and the second all the vertices. In this case, both centroid are located at $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ (1 options)
- One group includes two example and the second the other three
  i. The first group includes two adjacent vertices and the second all other three. For example, $G_1 = \{x_3, x_4\}$ with $\mu_1 = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$ and $G_2 = \{x_1, x_2, x_5\}$ with $\mu_2 = \begin{bmatrix} 0 \\ 2/3 \end{bmatrix}$. (4 options by symmetry)
  ii. The first group includes two vertices of opposing sides and the second include all other. In this case, both centroid are located at $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$. (2 options)

Notice that the partition in which one group include x_5 and one of the vertices is not a partition which the algorithm can converge to. Overall we got $1 + 4 + 1 + 4 + 2 = 12$ partitions.

(b) The partition which are not stable are those in which both centroid are located at $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$.