K-Means Clustering

Inputs: n objects (data points) $\{x_i\}_{i=1}^{n}$, and

K number of clusters desired

Outputs: · cluster ID for each data point, $z_i \in \{1, \dots K\}$.

· centroid of each cluster μ_K.

Idea: Alternatively optimize:

1) the assignment of data points to different clusters.
2) the centroids of the clusters.

objective:

$$\min_{z} \min_{\mu} \left\{ F(z, \mu) := \sum_{i=1}^{n} \| x_i - \mu_{z_i} \|^2 \right\}$$

where $z := \{z_i\}_{i=1}^{n} \in \{1, \dots K\}$

$$\mu := \{\mu_k\}_{k=1}^{K}.$$

Algorithm:

1) Initialize the centroids (or assignments) randomly;
2) Iterate until convergence:
   a) For all data points:

      Assign each data point (x_i) to a cluster that has the closest centroid:

$$\underset{\substack{\uparrow \\ \text{assigned centroid for data point } i.}}{z_i} = \underset{k=1,\dots,K}{\arg\min} \|x_i - \mu_k\|$$

   b) Update the centroids of all the clusters,
      For $k = 1$ to K:

$$\underset{\substack{\uparrow \\ \text{location of centroid } k}}{\mu_k} = \frac{1}{|S_k|} \sum_{i \in S_k} x_i \quad , \quad S_k = \{ i : z_i = k \}$$

K-Means as Optimization:

- K-Means can be viewed as a "coordinate descent" algorithm for optimizing an objective function of the centroids $\{\mu_K\}$ and assignments $\{z_i\}$.

Given: $\{x_i\}_{i=1}^{n}$

Want: $\underbrace{\{\mu_K\}_{K=1}^{K}}_{\mu}$ $\quad \underbrace{\{z_i\}_{i=1}^{n}}_{z}$

Define: $L(\mu, z) = \sum_{i=1}^{n} \|x_i - \mu_{z_i}\|^2$

Want: $\min_{\mu, z} L(\mu, z)$.

Mixed optimization: $\mu \in \mathbb{R}^{n \times k}$, $z \in \{1, \cdots, K\}^n$ $\{z_1 \dots z_n\}$

This will have lots of local minima.

Coordinate descent algorithm:

1) Initialize μ_0, z_0.
2) Repeat for t iterations:
   1) update z, with fixed μ,
      $$z_{t+1} = \arg\min_{z} L(\mu_t, z_t)$$
   2) update μ, with fixed z.
      $$\mu_{t+1} = \arg\min_{\mu} L(\mu_t, z_{t+1})$$

$$L(\mu, z) = \sum_{i=1}^{n} \|x_i - \mu_{z_i}\|^2$$

Find optimal $z_i \longleftarrow \underset{z_i}{\operatorname{argmin}} \|x_i - \mu_{z_i}\|^2$ // assignment step

Find optimal $\mu_k \longleftarrow \underset{\mu_k}{\operatorname{argmin}} \sum_{i \in S_k} \|x_i - \mu_k\|^2$

$S_k = \{ i : z_i = k \}$

$$= \sum_{i \in S_k} \|x_i\|^2 - 2x_i^T \mu_k + \|\mu_k\|^2$$

$$= \text{const} - 2\left(\sum_{i=1} x_i\right)^T \mu_k + |S_k| \|\mu_k\|^2$$

take derivative and set $= 0$.

$$\Rightarrow \mu_k = \frac{\sum_{i \in S_k} x_i}{|S_k|}$$ // centroid update step.

There is no guarantee in finding the global min.

$\Rightarrow$ the initializations of the centroids matter.

The algorithm can find the local optima.

Theorem:

$$(\mu_t, z_t) \Rightarrow L(\mu_{t+1}, z_{t+1}) \leq L(\mu_t, z_t)$$

proof: $L(\mu_t, z_t) \geq L(\mu_{t+1}, z_t)$

$$\geq L(\mu_{t+1}, z_{t+1})$$

The loss function can never increase with additional iteration.