Lecture: K - Means Algorithm

Scribe: Ramakrishnan P. Venkatachalam

The K-Means Algorithm

1. Initialize partitioning.
2. Compute induced mean vectors
3. Compute induced partitioning
4. If not converged, go to 2

K-means objective function:

$$\min \sum_{i=1}^{k} \sum_{x \epsilon C_i} ||x - m_i||_2{}^2$$

$$= \min \sum_{i=1}^{k} \sum_{x \epsilon C_i} (x - m_i)^T (x - m_i)$$

Theorem : The K-means objective function value never decreases.

$$\begin{aligned}
J^{(t)} &= \sum_{i} \sum_{x \epsilon C_i^{(t)}} ||x - m_i^{(t)}||^2 \\
\text{Let } ||x - m_j^{(t)}|| &= \min \, ||x - m_l^{(t)}||^2 \\
\text{Then } J^{(t)} &\geq \sum_{i} \sum_{x \epsilon C_i^{(t)}} ||x - m_j^{(t)}||^2 \\
&\geq \sum_{j} \sum_{x \epsilon C_i^{(t+1)}} ||x - m_j^{(t+1)}||^2 \\
&= J^{(t+1)}
\end{aligned}$$

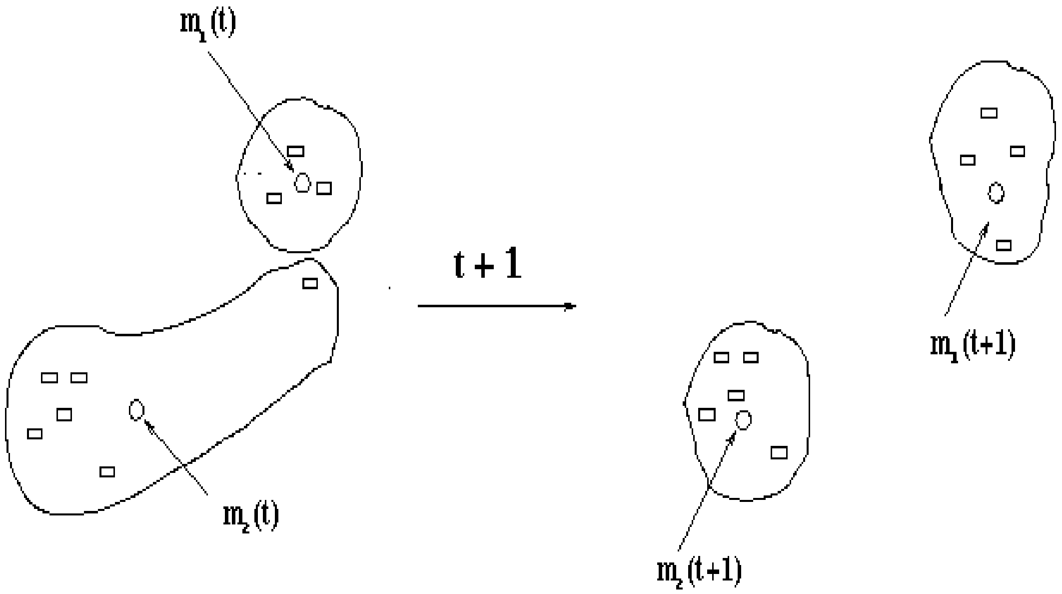

Figure 1: Iteration *t+1*

The k-means objective function and the algorithm results in the data points being partitioned by hyperplanes.

For example consider the case where k = 2.

$$x \in C_1 \quad \text{iff} \quad ||x - m_1||^2 \leq ||x - m_2||^2$$

The locus of points equidistant from m_1 and m_2 (i.e., the line passing perpendicular to the line joining m_1 and m_2 and passing through the mid point between m_1 and m_2) partitions the data points in such a fashion.

In d-dimensional space, the locus is the hyperplane normal to m_1 - m_2 and passing midway through the line joining m_1 - m_2. In general, when there are more than 2 clusters, k-means partitions the data set through as set of hyperplanes. The resulting partitions correspond to *Voronoi regions* or *Delaunay Triangulations.*

Duality of the k-means objective function

The k-means objective function can be thought of as $Trace(S_W)$ where

$$S_W = \sum_i \sum_{x \epsilon C_i} (x - m_i)(x - m_i)^T \, .$$

This is called the within-class scatter matrix.

Note: The trace of positive semi-definite matrices defines a norm on such matrices. Also $Trace(\Sigma xx^T) = \Sigma x^T x$

The total covariance matrix is given by

$$S_T = \sum_x (x-m)(x-m)^T,$$

where m is the mean. S_T is also known as the total scatter matrix.

It is possible to prove that $S_T = S_W + S_B$ where S_B is between class scatter matrix defined as follows:

$$S_B = \sum_i n_i (m_i - m)(m_i - m)^T$$

Since $Trace(S_T) = Trace(S_W) + Trace(S_B)$ and $Trace(S_T)$ is fixed, we can view k-means objective function as

$$\max \sum_i n_i ||m_i - m||^2$$

$$= \max \frac{1}{2n} \sum_i \sum_j n_i n_j ||m_i - m_j||^2$$

i.e., k-means can be though of as maximizing the pairwise distance between the mean vectors.

Proof that $S_T = S_W + S_B$

$$\begin{aligned}
&Define \\
S_T(a) &= \sum_x (x-a)(x-a)^T, \\
&\quad a\ is\ any\ vector \\
&= \sum_i \sum_{x \epsilon C_i} (x - a + m_i - m_i)(x - a + m_i - m_i)^T \\
&= \sum_i \sum_{x \epsilon C_i} ((x - m_i) + (m_i - a))((x - m_i) + (m_i - a)^T \\
&= \sum_i \sum_{x \epsilon C_i} (x - m_i)(x - m_i)^T + \sum_i \sum_{x \epsilon C_i} (m_i - a)(x - m_i)^T \\
&\quad + \sum_i \sum_{x \epsilon C_i} (m_i - a)(x - m_i)^T + \sum_i \sum_{x \epsilon C_i} (m_i - a)(m_i - a)^T \\
S_T(a) &= \sum_i \sum_{x \epsilon C_i} (x - m_i)(x - m_i)^T + \sum_i n_i (m_i - a)(m_i - a)^T, \\
&\quad since\ the\ second\ and\ third\ terms\ in\ the\ previous\ equation\ are\ zero. \\
&= S_W + S_B(a)
\end{aligned}$$

The above equation is true for an a. In particular for a = m. So we are done.

We can also show that

$$
\begin{aligned}
S_T &= \frac{1}{2n}\sum_i \sum_j (x_i - x_j)(x_i - x_j)^T \\
S_W &= \sum_i \frac{1}{2n_i} \sum_{x \in C_i} \sum_{y \in C_i} (x-y)(x-y)^T \\
&\textit{and} \\
S_B &= \frac{1}{2n}\sum_i \sum_j n_i n_j (m_i - m_j)(m_i - m_j)^T
\end{aligned}
$$

Since $Trace(S_W)$ = k-means objective function, k-means minimizes the average pairwise sum-of-square of distances between points within the clusters.

Proof that $S_T = \frac{1}{2n}\sum_i \sum_j (x_i - x_j)(x_i - x_j)^T$

$$
\begin{aligned}
Now\ S_T &= \sum_x (x-m)(x-m)^T \\
&= \sum_x xx^T + \sum_x (mx^T + xm^T) + nmm^T \\
Pairwise\ scatter\ S_P &= \sum_i \sum_j (x_i - x_j)(x_i - x_j)^T \\
&= \sum_i \sum_j (x_i x_i^T - x_j x_i^T - x_i x_j^T + x_j x_j^T) \\
&= 2n \sum_x xx^T - \sum_j x_j (\sum_i x_i{}^T) - \sum_i x_i (\sum_j x_j^T) \\
&= 2n \sum_x xx^T - \sum_j x_j nmm^T - \sum_i x_i nmm^T \\
&= 2n \sum_x xx^T - 2n^2 mm^T \\
\therefore S_T &= \frac{S_P}{2n}
\end{aligned}
$$