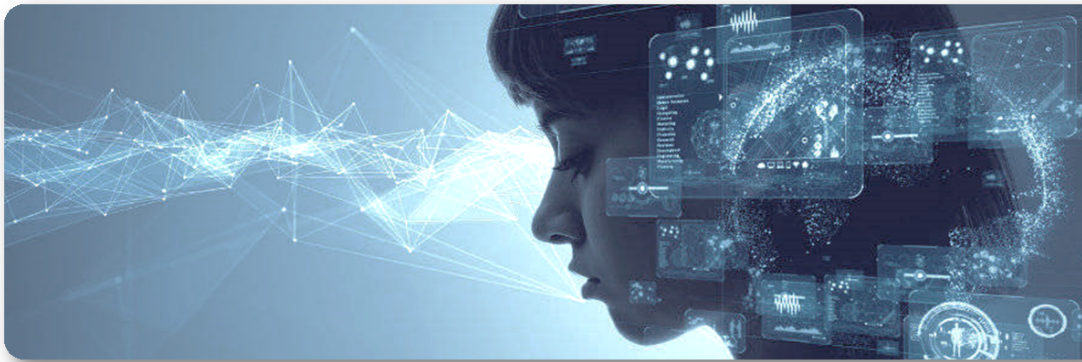

COMP 2211 Exploring Artificial Intelligence

K-Means Clustering - Why the described algorithm works? Also, will it converge?

Prof. Song Guo, Dr. Desmond Tsoi & Dr. Huiru Xiao

Department of Computer Science & Engineering
The Hong Kong University of Science and Technology, Hong Kong SAR, China

Why?

- Recall that a clustering algorithm must return both a clustering and a centre for each cluster.
- The following proves the Sum of Square Error (SSE) is minimized when the centre associated with each cluster is the mean (or centroid) of the set of points assigned to that cluster.
- Consider the points $\mathbf{x_1}$, $\mathbf{x_2}$, ..., $\mathbf{x_m}$, where $m \geq 1$, and for $i \in \{1, 2, \ldots, m\}$, $\mathbf{x_i}$ is in d-dimensional.
- Let $\bar{\mathbf{x}} = \frac{\mathbf{1}}{\mathbf{m}} \sum_{\mathbf{i=1}}^{\mathbf{m}} \mathbf{x_i}$ be the mean of these points, and let $\mathbf{x}$ is in d-dimensional be an arbitrary point in the same d-dimensional space.

What to Prove?

- We want to prove the following:

$$\sum_{i=1}^{m} ||\mathbf{x_i} - \mathbf{x}||^2 \geq \sum_{i=1}^{m} ||\mathbf{x_i} - \overline{\mathbf{x}}||^2$$

Note

$$\overline{\mathbf{x}} = \frac{(\mathbf{x_1} + \mathbf{x_2} + \ldots + \mathbf{x_m})}{m}$$

$$\text{So, } m\overline{\mathbf{x}} = (\mathbf{x_1} + \mathbf{x_2} + \ldots + \mathbf{x_m})$$

Proof

$$
\begin{aligned}
\sum_{i=1}^{m} ||\mathbf{x_i} - \mathbf{x}|| &= \sum_{i=1}^{m} ||(\mathbf{x_i} - \bar{\mathbf{x}}) + (\bar{\mathbf{x}} - \mathbf{x})||^2 \\
&= \sum_{i=1}^{m} (||\mathbf{x_i} - \bar{\mathbf{x}}||^2 + ||\bar{\mathbf{x}} - \mathbf{x}||^2 + 2(\mathbf{x_i} - \bar{\mathbf{x}}) \cdot (\bar{\mathbf{x}} - \mathbf{x})) \\
&= \sum_{i=1}^{m} ||\mathbf{x_i} - \bar{\mathbf{x}}||^2 + \sum_{i=1}^{m} ||\bar{\mathbf{x}} - \mathbf{x}||^2 + 2 \sum_{i=1}^{m} (\mathbf{x_i} \cdot \bar{\mathbf{x}} - \mathbf{x_i} \cdot \mathbf{x} - \bar{\mathbf{x}} \cdot \bar{\mathbf{x}} + \bar{\mathbf{x}} \cdot \mathbf{x}) \\
&= \sum_{i=1}^{m} ||\mathbf{x_i} - \bar{\mathbf{x}}||^2 + m||\bar{\mathbf{x}} - \mathbf{x}||^2 + 2(m\bar{\mathbf{x}} \cdot \bar{\mathbf{x}} - m\bar{\mathbf{x}} \cdot \mathbf{x} - m\bar{\mathbf{x}} \cdot \bar{\mathbf{x}} + m\bar{\mathbf{x}} \cdot \mathbf{x}) \\
&= \sum_{i=1}^{m} ||\mathbf{x_i} - \bar{\mathbf{x}}||^2 + m||\bar{\mathbf{x}} - \mathbf{x}||^2 \\
&\geq \sum_{i=1}^{m} ||\mathbf{x_i} - \bar{\mathbf{x}}||^2
\end{aligned}
$$

Will K-Means Clustering algorithm converge?

- K-Means clustering algorithm converges to a local optimum.
- To prove convergence of the K-Means algorithm, we show that the loss function is guaranteed to decrease monotonically in each iteration until convergence for the assignment step and for the refitting step. Since the loss function is non-negative, the algorithm will eventually converge when the loss function reaches its local minimum.
- Let $z = (z_1, \ldots, z_n)$ denote the cluster assignments for the n points.

Assignment Step

- We can write down the original loss function $L(\mu)$ as follows:

$$L(\mu, z) = \sum_{i=1}^{n} ||x_i - \mu_{z_i}||_2^2$$

  where $|| \cdot ||_2^2$ denote square of L2-norm.

- Let's consider a data point x_i, and let z_i be the assignment from the previous iteration and z_i^* be the new assignment obtained as:

$$z_i^* \in argmin_{j \in \{1,\dots,k\}} ||x_i - \mu_j||_2^2$$

- Let z^* denote the new cluster assignments for all the n points. The change in loss function after this assignment step is then given by:

$$L(\mu, z^*) - L(\mu, z) = \sum_{i=1}^{n} (||x_i - \mu_{z_i^*}||_2^2 - ||x_i - \mu_{z_i}||_2^2) \leq 0$$

  The inequality holds by the rule z_i^* is determined, i.e. to assign x_i to the nearest cluster.

Refitting Step

- We can write down the original loss function $L(\mu)$ as follows:

$$L(\mu, z) = \sum_{j=1}^{k} \left(\sum_{i:z_i=j} ||x_i - \mu_j||_2^2 \right)$$

- Let's consider the j^{th} cluster, and let μ_j be the cluster center from the previous iteration and μ_j^* be the new cluster center obtained as:

$$\mu_j^* = \frac{1}{|\{i : z_i = j\}|} \sum_{i:z_i=j} x_i$$

- Let μ^* denote the new cluster centers for all the k clusters. The change in loss function after this refitting step is then given by:

$$L(\mu^*, z) - L(\mu, z) = \sum_{j=1}^{k} \left(\left(\sum_{i:z_i=j} ||x_i - \mu_j^*||_2^2 \right) - \left(\sum_{i:z_i=j} ||x_i - \mu_j||_2^2 \right) \right) \leq 0$$

  This inequality holds because the update rule of μ_j^* essentially minimizes this quantity.

That's all!

Any questions?