K-means Algorithm Mathematics base

Problem:

Given a data set $\{x_1, \ldots, x_n\} \in \mathbb{R}^{\mathrm{d}}$ and a integer number K<=N. Our goal is to partition the data set into K clusters.

Let $\{u_i\} | u_i \in \mathbb{R}^{\mathrm{d}}, i = 1, .., K$ be the set of center point of each cluster. Our goad is then to find an assignment of data points to clusters such that sum of the squares of the distances of each data point to its closest vector u_k is minimum.

Find min:

$||x_i - u_k||_2^2$

For each data point x_i, we introduce a set $\{y_{ij} \in \{0, 1\}\}$ where j = 1,...,K. If data point x_i is assigned to cluster k then $y_{ik} = 1, and\, y_{ij} = 0\ for\ i \neq j$. Now, we can define an objective function:

$\mathcal{L}(\mathbf{Y}, \mathbf{U}) = \sum_{i=1}^{N} \sum_{j=1}^{K} y_{ij} \|\mathbf{x}_i - \mathbf{u}_j\|_2^2$

We can minimize this function through an iterative procedure in which each iteration involves two successive steps corresponding to successive optimizations with respect to the y_{ij} and u_k.

- fixed U, find Y:

  $\mathbf{y}_i = \arg\min_{\mathbf{y}_i} \sum_{j=1}^{K} y_{ij} \|\mathbf{x}_i - \mathbf{u}_j\|_2^2 \quad (3)$

  $\text{subject to: } y_{ij} \in \{0, 1\}\ \forall j; \quad \sum_{j=1}^{K} y_{ij} = 1$

  $<=> j = \arg\min_j \|\mathbf{x}_i - \mathbf{u}_j\|_2^2$

  In other words, we simply assign point x_i to the closest cluster center.

- fixed Y, find U:

  $\mathbf{u}_j = \arg\min_{\mathbf{u}_j} \sum_{i=1}^{N} y_{ij} \|\mathbf{x}_i - \mathbf{u}_j\|_2^2.$

  Objective function is a quadratic function of u_j so it can be minimized by setting its derivative to 0:

  $\frac{\partial l(\mathbf{u}_j)}{\partial \mathbf{u}_j} = 2 \sum_{i=1}^{N} y_{ij} (\mathbf{u}_j - \mathbf{x}_i) = 0$

  $\Rightarrow \mathbf{u}_j = \frac{\sum_{i=1}^{N} y_{ij} \mathbf{x}_i}{\sum_{i=1}^{N} y_{ij}}$

  The denominator in this expression is equal to the number of points assigned to cluster k, the numerator is sum of all points of cluster j. So this result has a simple interpretation, namely set u_j equal to the mean of all of the data points x_i assigned to cluster j.

  Because each phase reduces the value of the objective function, convergence of the algorithm is assured. However, it may converge to a local rather than global minimum.