LECTURE 28 - K-MEANS & FCM

1. K-Means Clustering

Algorithm 1 K-Means Algorithm

Set number of clusters, M
Initialize cluster centers
repeat
    for i = 1 to N **do**
        Determine the closest representative, Θ_j, for $\mathbf{x}_i$
        Set label for data point i to j
    end for
    for j = 1 to M **do**
        Update cluster representative Θ_j to the mean of the points with cluster label j
    end for
until Change in cluster centers is small

(1) *Run code example*
(2) Objective Function for K-means:

$$J(\Theta, U) = \sum_{i=1}^{N} \sum_{j=1}^{m} u_{ij} \left\| \mathbf{x}_i - \theta_j \right\|^2 \tag{1}$$

- *How would you optimize the objective? Does the pseudo-code above match the obj. func. optimization?*
- *Does the K-means algorithm find the* globally optimal *solution (i.e., the cluster centers and assignments that globally minimize the objective function)?*
- *Does the K-means algorithm make any assumptions on cluster shape?*
- *Given a data set with an unknown number of clusters, can you come up with a strategy for determining the "right" number of clusters?*

2. Fuzzy C-Means Clustering

Motivation for the fuzzy memberships examples: point mid-way between clusters, overlapping clusters

So, we want soft clustering (as opposed to the hard/crisp clustering given by sequential or k-means). We can do that by defining an objective function with membership values.

The FCM objective function is:

$$J = \sum_{i=1}^{N} \sum_{j=1}^{M} u_{ij}^{m} d(\mathbf{x}_i, C_j) \quad \text{s.t. } 0 \leq u_{ij} \leq 1; \sum_{j=1}^{M} u_{ij} = 1$$

Suppose $d(\mathbf{x}_i, C_j) = \|\mathbf{x}_i - \theta_j\|_2^2$ where θ_j is the mean for cluster j, how would you minimize this objective? How do you incorporate the constraints on the membership values?

(1) *Solve for the update equation for the cluster representatives:*

$$J = \sum_{i=1}^{N} \sum_{j=1}^{M} u_{ij}^{m} \|\mathbf{x}_i - \theta_j\|_2^2 \tag{2}$$

$$\frac{\partial J}{\partial \theta_j} = \sum_{i=1}^{N} -2u_{ij}^{m} (\mathbf{x}_i - \theta_j) \tag{3}$$

$$0 = \sum_{i=1}^{N} \left(-2u_{ij}^{m}\mathbf{x}_i + 2u_{ij}^{m}\theta_j\right) \tag{4}$$

$$\sum_{i=1}^{N} u_{ij}^{m}\mathbf{x}_i = \theta_j \sum_{i=1}^{N} u_{ij}^{m} \tag{5}$$

$$\theta_j = \frac{\sum_{i=1}^{N} u_{ij}^{m}\mathbf{x}_i}{\sum_{i=1}^{N} u_{ij}^{m}} \tag{6}$$

(2) *Solve for the update equation for the membership values:*

$$L = \sum_{i=1}^{N}\sum_{j=1}^{M} u_{ij}^{m} \left\|\mathbf{x}_i - \theta_j\right\|_2^2 - \sum_{i=1}^{N}\lambda_i\left(\sum_{j=1}^{M} u_{ij} - 1\right) \tag{7}$$

$$\frac{\partial J}{\partial u_{ij}} = m u_{ij}^{(m-1)} \left\|\mathbf{x}_i - \theta_j\right\|_2^2 - \lambda_i = 0 \tag{8}$$

$$u_{ij} = \left(\frac{\lambda_i}{m\left\|\mathbf{x}_i - \theta_j\right\|_2^2}\right)^{\frac{1}{(m-1)}} \tag{9}$$

$$\sum_{j=1}^{M} u_{ij} = 1 \tag{10}$$

$$1 = \sum_{j=1}^{M}\left(\frac{\lambda_i^{\frac{1}{(m-1)}}}{\left(m\left\|\mathbf{x}_i - \theta_j\right\|_2^2\right)^{\frac{1}{(m-1)}}}\right) \tag{11}$$

$$1 = \lambda_i^{\frac{1}{(m-1)}}\sum_{j=1}^{M}\left(\frac{1}{\left(m\left\|\mathbf{x}_i - \theta_j\right\|_2^2\right)^{\frac{1}{(m-1)}}}\right) \tag{12}$$

$$\lambda_i^{\frac{1}{(m-1)}} = \left(\sum_{k=1}^{M}\left(\frac{1}{\left(m\left\|\mathbf{x}_i - \theta_k\right\|_2^2\right)^{\frac{1}{(m-1)}}}\right)\right)^{-1} \tag{13}$$

$$u_{ij} = \left(\sum_{k=1}^{M}\left(\frac{1}{\left(m\left\|\mathbf{x}_i - \theta_k\right\|_2^2\right)^{\frac{1}{(m-1)}}}\right)\right)^{-1}\left(\frac{1}{m\left\|\mathbf{x}_i - \theta_j\right\|_2^2}\right)^{\frac{1}{(m-1)}} \tag{14}$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{M}\frac{1}{\left(m\left\|\mathbf{x}_i - \theta_k\right\|_2^2\right)^{\frac{1}{(m-1)}}}}\left(\frac{1}{m\left\|\mathbf{x}_i - \theta_j\right\|_2^2}\right)^{\frac{1}{(m-1)}} \tag{15}$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{M}\left(\frac{\left\|\mathbf{x}_i - \theta_j\right\|_2^2}{\left\|\mathbf{x}_i - \theta_k\right\|_2^2}\right)^{\frac{1}{m-1}}} \tag{16}$$

(1) *What constraints are there on the value of the fuzzifier m? Why?*
(2) *How do results change as m approaches 1? 2? 3? How do results changes with changes in m? Can m get too big? too small?*
(3) *Does FCM find the globally optimal solution? Why or why not? What sort of optimization strategy is being used in the FCM algorithm?*

Algorithm 2 Fuzzy C-Means Algorithm

Set number of clusters, M, and q
Initialize cluster centers
repeat
  for i = 1 to N **do**
    for j = 1 to M **do**
      if $d(\mathbf{x}_i, C_j) \neq 0$ **then**
        $u_{ij} \leftarrow \frac{1}{\sum_{k=1}^{M}\left(\frac{d(\mathbf{x}_i, C_j)}{d(\mathbf{x}_i, C_k)}\right)^{\frac{1}{q}}}$
      else
        $u_{ij} = 1$
      end if
    end for
  end for
  for j = 1 to M **do**
    Update cluster representatives
  end for
until Change in cluster centers is small

(4) *What FCM membership values would a point that is located exactly halfway between two cluster centers (when clustering into two clusters) obtain?*
(5) *What FCM membership values would an outlier data point recieve? An outlier data point would be one that is far from all cluster representatives.*
(6) *What may be some advantages of FCM over K-Means? disadvantages?*