DraftTesiMazza

February 2025

1 MIP Formulation for K-Means

The formulation is for each day d and session s. This dependency is avoided in the notation.

1.1 Sets

- $\mathcal{P} := \{1, ..., N\}$: set of patients of day d and session s;
- $\mathcal{R} := \{1, ..., R\}$: set of requests of day d and session s.

1.2 Parameters

- $\tau_{ij} \geq 1$ = time in minutes between patient i and j;
- $K \in \mathbb{N}, K > 1$ number of clusters;
- $w_i \in [0,1] = \frac{|\{j \ | \ p_j = i \quad \forall j \in \mathcal{R}\}|}{R}$

1.3 Variables

- $x_{ik} \in \{0,1\} = \mathbb{1}\{\text{patient i is associated to cluster k}\}$;
- $y_{ik} \in \{0,1\} = \mathbb{1}\{\text{patient i is the centroid of cluster k}\}$;

1.4 Formulation

$$\min_{x,y} \frac{1}{K} \sum_k \sum_i d_{ik}$$

$$\sum_k x_{ik} = 1 \qquad \forall i \in \mathcal{P} \tag{1}$$

$$\sum_i y_{ik} = 1 \qquad \forall k = 1, .., K \tag{2}$$

$$\sum_i \sum_k y_{ik} = K \tag{3}$$

$$y_{ik} \leq x_{ik} \qquad \forall i \in \mathcal{P},\ \forall k = 1, .., K \tag{4}$$

$$d_{ik} = \sum_j y_{jk}\tau_{ij} \qquad \forall i \in \mathcal{P},\ \forall k = 1, .., K \tag{5}$$

$$d_{ik} = \sum_j y_{jk}\tau_{ij} w_j w_i? \qquad \forall i \in \mathcal{P},\ \forall k = 1, .., K \tag{6}$$

2 Minimax Clustering

K-Means approach works well for compact, spherical clusters, but it can struggle with outliers or cases where the maximum distance within a cluster is too large, leading to uneven groupings.

As an alternative, Minimax Clustering addresses this limitation by minimizing the maximum intra-cluster distance, also known as the cluster diameter, which is the largest distance between any two points within the same cluster. This ensures more compact and balanced clusters, reducing the impact of widely spread data points.

2.1 MIP Formulation

With the same variables and parameters, define:

- $C_k := \{i \mid x_{ik} = 1, \forall i \in \mathcal{P}\}$ cluster k;
- $d_k = \max_{i,j \in C_k} \tau_{ij}$ the diameter of cluster k;
- $M = \max_{i,j \in \mathcal{P}} \tau_{ij}$

$$\min_x \frac{1}{K} \sum_k d_k$$

$$\sum_k x_{ik} = 1 \qquad \forall i \in \mathcal{P} \tag{7}$$

$$\sum_i x_{ik} \geq 1 \qquad \forall k = 1, .., K \tag{8}$$

$$d_k \geq \tau_{ij} - M(2 - x_{ik} - x_{jk}) \qquad \forall i, j \in \mathcal{P},\ \forall k = 1, .., K \tag{9}$$

3 Open Questions

- How to weight the distance?;