DraftTesiMazza

February 2025

1 K-Medoids

K-medoids is a clustering algorithm that selects actual data points as cluster centers (medoids) instead of computing centroids like k-means. It minimizes the total dissimilarity (e.g., distance or travel time) between points and their assigned medoid, making it more robust to outliers and suitable for arbitrary distance metrics.

K-medoids is more suitable than k-means when working with distance metrics like travel times, which may not be symmetric, since avoids the need for computing a "mean" point in a non-Euclidean space.

1.1 MIP Formulation

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

1.2 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.

1.3 Parameters

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

1.4 Variables

- $x_{ij} \in \{0, 1\} = \mathbb{1}\{\text{patient i is associated to medoid j}\}$;
- $y_i \in \{0, 1\} = \mathbb{1}\{\text{patient i is a medoid}\}$;

1.5 Formulation

$$\min_{x,y} \sum_i \sum_j \tau_{ij} x_{ij}$$

$$\sum_j x_{ij} = 1 \qquad \forall i \in \mathcal{P} \tag{1}$$

$$\sum_i y_i = K \tag{2}$$

$$x_{ij} \leq y_j \qquad \forall i, j \in \mathcal{P} \tag{3}$$

2 Minimax Clustering

Questo va ripensato

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;

- $M = \max_{i,j \in \mathcal{P}} \tau_{ij}$

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

$$\sum_k x_{ik} = 1 \qquad \forall i \in \mathcal{P} \tag{4}$$

$$\sum_i x_{ik} \geq 1 \qquad \forall k = 1, .., K \tag{5}$$

$$d_k \geq \tau_{ij} - M(2 - x_{ik} - x_{jk}) \qquad \forall i, j \in \mathcal{P},\ \forall k = 1, .., K \tag{6}$$

3 Open Questions

- How to weight the distance?;