An approach to cluster separability in a partition

K. Sabo[a,*], R. Scitovski[a]

[a] *Department of Mathematics, University of Osijek, Trg Ljudevita Gaja 6, HR-31000 Osijek, Croatia*

Abstract

In this paper, we consider the problem of cluster separability in a minimum distance partition based on the squared Euclidean distance. We give a characterization of a well-separated partition and provide an operational criterion that gives the possibility to measure the quality of cluster separability in a partition. Especially, the analysis of cluster separability in a partition is illustrated by implementation of the k-means algorithm.

Keywords: clustering, data mining, cluster separability, separability balls

2000 MSC: 62H30, 68T10, 90C26, 90C27, 91C20, 47N10

*Corresponding author

Email addresses: `ksabo@mathos.hr` (K. Sabo), `scitowsk@mathos.hr` (R. Scitovski)

1. Introduction

Clustering or grouping a set of data points into conceptually meaningful clusters is a well-studied problem in recent literature [2, 3, 9, 11, 19, 21, 23, 28], and it has practical importance in a wide variety of applications such as computer vision, signal-image-video analysis, multimedia, networks, biology, medicine, geology, psychology, business, politics and other social sciences.

Let $I = \{1, \dots, m\}$ and $J = \{1, \dots, k\}$. A partition of the set $\mathcal{A} = \{a_i \in \mathbb{R}^n : i \in I\}$ into k disjoint subsets $\pi_1, \dots, \pi_k$, $1 \le k \le m$, such that

$$\bigcup_{i=1}^{k} \pi_i = \mathcal{A}, \qquad \pi_r \cap \pi_s = \emptyset, \quad r \neq s, \qquad |\pi_j| \ge 1, \quad \forall r, s, j \in J, \tag{1}$$

will be denoted by $\Pi = \{\pi_1, \dots, \pi_k\}$ and the set of all such partitions by $\mathcal{P}(\mathcal{A}, k)$. The elements $\pi_1, \dots, \pi_k$ of the partition Π are called *clusters in* $\mathbb{R}^n$.

Any function $d\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$, $\mathbb{R}_+ := [0, +\infty\rangle$, with the following property

$$(\forall (x, y) \in \mathbb{R}^n \times \mathbb{R}^n) \quad d(x, y) \ge 0 \quad \text{and} \quad d(x, y) = 0 \Leftrightarrow x = y,$$

is called a distance-like function (see, e.g., [11, 28]). Let $d\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$, be a distance-like function. Then for each cluster $\pi_j \in \Pi$ its center c_j is defined by

$$c_j = c(\pi_j) := \operatorname*{argmin}_{x \in \operatorname{conv} \pi_j} \sum_{a_i \in \pi_j} d(x, a_i), \tag{2}$$

where $\operatorname{conv} \pi_j$ denotes the convex hull of the cluster π_j. It is said that the partition $\Pi^\star \in \mathcal{P}(\mathcal{A}, k)$ is a globally optimal k-partition if

$$\Pi^\star = \operatorname*{argmin}_{\Pi \in \mathcal{P}(\mathcal{A}, k)} \mathcal{F}(\Pi), \qquad \mathcal{F}(\Pi) = \sum_{j=1}^{k} \sum_{a_i \in \pi_j} d(c_j, a_i), \tag{3}$$

where $\mathcal{F}\colon \mathcal{P}(\mathcal{A}, k) \to \mathbb{R}_+$ is the objective function.

Conversely, for a given set of different points $z_1, \dots, z_k \in \mathbb{R}^n$, by applying the *minimum distance principle* (see, e.g., [11, 25]), one can define the partition $\Pi = \{\pi(z_1), \dots, \pi(z_k)\}$,

$$\pi(z_j) = \{a \in \mathcal{A} : d(z_j, a) \le d(z_s, a), \forall s = 1, \dots, k\}, \qquad j \in J, \tag{4}$$

where a tie-breaker rule is needed in case of equality.

Therefore, the problem of finding an optimal partition of the set $\mathcal{A}$ can be reduced to the following optimization problem:

$$\underset{z_1,\ldots,z_k\in\mathbb{R}^n}{\operatorname{argmin}} F(z_1,\ldots,z_k), \qquad F(z_1,\ldots,z_k) = \sum_{i=1}^{m} \min_{1\le j\le k} d(z_j, a_i). \tag{5}$$

Optimization problems (3) and (5) are equivalent [25]. Global optimization problem (5) can also be found in the literature as a *center-based clustering problem* [9, 13, 28]. If the *squared Euclidean distance* $d\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$, $d(x, y) = \|x - y\|^2$ is used, the function F from (5) becomes a standard $k-$*means* objective function. The objective function $F\colon \mathbb{R}^{kn} \to \mathbb{R}_+$ defined by (5) can have a large number of independent variables (the number of clusters in the partition multiplied by the dimension of data points: $k \cdot n$), it does not have to be either convex or differentiable and usually it has several local minima. Hence, this becomes a complex global optimization problem.

Furthermore, suppose that $\mathcal{A} \subset \mathbb{R}^n = \{(x_1,\ldots,x_n)\colon x_i \in \mathbb{R}\}$ is a given set. By using the *squared Euclidean distance* $d\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_+$, $d(x, y) = \|x - y\|^2 = \langle x - y, x - y\rangle$, where $\langle\cdot,\cdot\rangle$ is the standard inner product, we analyze internal separability of some partition Π of the set of data points $\mathcal{A}$, i.e., we consider the following problem:

> Let $\mathcal{A} \subset \mathbb{R}^n$ be a set, d the squared Euclidean distance and $z_1,\ldots,z_k \in \mathbb{R}^n$ a set of mutually different points (*assignment points*) that determine the partition $\Pi = \{\pi(z_1),\ldots,\pi(z_k)\}$, where $\pi(z_j)$ are given by (4). The question is: *How can the assignment points be changed such that the partition* Π *remains unchanged?*

Especially, an open ball $B(\delta) = \{u \in \mathbb{R}^n\colon \|u\| < \delta\}$ of radius $\delta > 0$ is searched for, such that for an arbitrary set of assignment points $\{\zeta_1,\ldots,\zeta_k \in \mathbb{R}^n\colon \zeta_j \in z_j + B(\delta)\}$ the clusters $\pi(\zeta_j)$ and $\pi(z_j)$ are equal for all $j \in J$. The ball $B(\delta)$ is said to be a *separability ball of the partition* Π and the corresponding balls

$$z_j + B(\delta) := \{z_j + u\colon u \in B(\delta)\}, \quad j \in J,$$

will be called *separability balls associated with assignment points* $z_1,\ldots,z_k$.

Note that in this way separability balls for all clusters have the same radius δ. The problem could also be formulated such that separability balls are searched for each cluster separately.

There is a rich literature considering similar problems. Some of them will be discussed in detail in the next section, after the term cluster separability in a partition is defined and a characterization of a well-separated partition is given. The problem is first considered for the one-dimensional case, and then in detail for the n-dimensional case. In Section 3, cluster separability in a partition is illustrated by the implementation of the k-means algorithm. Finally, some conclusions are given in Section 4.

2. Cluster separability in a partition

Let $1 \le k \le m$, $I = \{1, \dots, m\}$, $J = \{1, \dots, k\}$, and let $\mathcal{A} = \{a_i \in \mathbb{R}^n \colon i \in I\}$ be a given data set in $\mathbb{R}^n$. By using the squared Euclidean distance, for a given set of assignment points $z_1, \dots, z_k \in \mathbb{R}^n$, according to the minimum distance principle, there is a partition $\Pi = \{\pi(z_1), \dots, \pi(z_k)\}$ made up of clusters

$$\pi(z_j) = \{a \in \mathcal{A} \colon \|z_j - a\| \le \|z_s - a\|,\ s \in J\}, \quad j \in J. \tag{6}$$

Note that each cluster $\pi(z_j)$ depends on the neighboring clusters, and notation $\pi(z_j)$ implies that cluster $\pi(z_j)$ is associated to the center z_j. It is well-known (see, e.g., [11]) that it may happen that some of the clusters are empty sets or that some elements $a \in \mathcal{A}$ appear on the border of two or more clusters $\pi(z_1), \dots, \pi(z_k)$ determined by assignment points $z_1, \dots, z_k$ (see e.g., [22]). In the latter case, such an element is associated only to one of the clusters whose boundary it lies on. Also, note that equation (6) expresses that fact that the cluster $\pi(z_j)$ is the intersection of the Voronoi cell (see, e.g. [1, 15]) $\{x \in \mathbb{R}^n : \|x - z_j\| \le \|x - z_s\|\ \forall s \ne j\}$ with the dataset $\mathcal{A}$.

Example 1. *[25] Let $n = k = 2$. All data points $a \in \mathcal{A} \subset \mathbb{R}^2$ lying on the perpendicular bisector of the line segment $\overline{z_1 z_2}$,*

$$\sigma[z_1, z_2] = \{a \in \mathbb{R}^2 \colon \langle z_2 - z_1, a - \tfrac{1}{2}(z_1 + z_2)\rangle = 0\},$$

passing through the midpoint of that line segment are placed equidistant from the points z_1 and z_2. If a data point lies on the border between the two clusters, it can be associated either to the first or to the second cluster.

First, we define the term *well-separated partition* of two clusters in $\mathbb{R}^n$ (see Fig. 1), and after that the definition is generalized for partitions with $1 \leq k \leq m$ clusters.

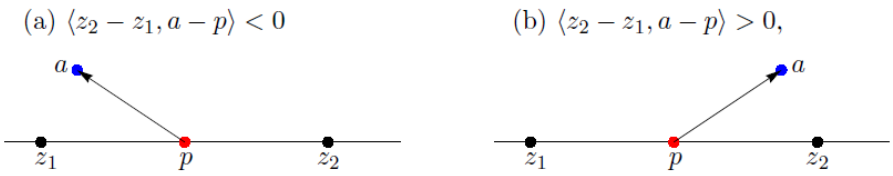


Figure 1: The minimum distance principle

Definition 1. Let $\mathcal{A} \subset \mathbb{R}^n$ be a data set and $z_1, z_2 \in \mathbb{R}^n$ two different assignment points. It is said that the partition $\Pi = \{\pi(z_1), \pi(z_2)\}$ consisting of two clusters and defined according to the minimum distance principle (4) *is a well-separated partition* if $\pi(z_1), \pi(z_2) \neq \emptyset$, and if for all $a \in \mathcal{A}$ the following holds

$$\langle z_2 - z_1, a - p\rangle \neq 0, \quad p = \tfrac{1}{2}(z_1 + z_2). \tag{7}$$

Geometrically, inequality (7) means that there is no data point $a \in \mathcal{A}$ which lies on the bisecting hyperplane. In addition, the following holds:

$$\{a \in \mathcal{A} \colon \langle z_2 - z_1, a - p\rangle < 0\} = \pi(z_1) \qquad \text{(Fig. 1a)}, \tag{8}$$

$$\{a \in \mathcal{A} \colon \langle z_2 - z_1, a - p\rangle > 0\} = \pi(z_2) \qquad \text{(Fig. 1b)}. \tag{9}$$

Note that according to Definition 1, any $a \in \mathcal{A}$ belongs to the cluster $\pi(z_1)$ if the distance from a to the assignment point z_1 is less than the distance to the assignment point z_2. This will occur if $\angle(z_2 - z_1, a - p)$ is an obtuse angle (Fig. 1a). The point $a \in \mathcal{A}$ belongs to the cluster $\pi(z_2)$ if $\angle(z_2 - z_1, a - p)$ is an acute angle (Fig. 1b). If for some $a_{i_0} \in \mathcal{A}$, $\langle z_2 - z_1, a_{i_0} - p\rangle = 0$ (a_{i_0} lies on the border between clusters $\pi(z_1)$ and $\pi(z_2)$), it is said that the partition $\Pi = \{\pi(z_1), \pi(z_2)\}$ is not well separated.

The following definition is a natural generalization of Definition 1.

Definition 2. Let $\mathcal{A} \subset \mathbb{R}^n$ be a data set and $z_1, \ldots, z_k \in \mathbb{R}^n$ mutually different assignment points. It is said that the partition $\Pi = \{\pi(z_1), \ldots, \pi(z_k)\}$ consisting of k clusters and defined according to the minimum distance principle *is a well-separated partition* if $\pi(z_j) \neq \emptyset$, $j \in J$, and if for each pair $1 \leq j < s \leq k$ and for all $a \in \pi(z_j) \cup \pi(z_s)$ the following holds

$$\langle z_j - z_s, a - p(z_j, z_s)\rangle \neq 0, \quad \text{where } p(z_j, z_s) = \tfrac{1}{2}(z_j + z_s).$$

Geometrically, this inequality means that for each pair of indices $1 \leq j < s \leq k$ no data point $a \in \pi(z_j) \cup \pi(z_s)$ lies on the bisecting hyperplane between z_j and z_s. Thereby

$$\{a \in \mathcal{A} \colon \langle z_j - z_s, a - p(z_j, z_s)\rangle < 0, \forall s \in J \setminus \{j\}\} = \pi(z_j), \quad j \in J.$$

In [27], the term *stable partition* is defined as a partition unchanged by an iteration of k-means and its properties are given. Particularly, it has been shown that if $\Pi = \{\pi_1, \ldots, \pi_k\}$ is a stable partition; then for all $a \in \mathcal{A}$ there is a unique nearest assignment point. From this statement it consequently follows that every *stable partition* is necessarily a *well-separated partition* in accordance with Definition 1 and Definition 2. Obviously, the converse is not true, i.e., there exists a well-separated partition that is not stable.

Similarly, in the literature (see, e.g., [4, 5, 8, 12, 14, 16, 24]), cluster stability in a partition is usually considered as a property of cluster elements, that small perturbations in the data do not significantly influence to which cluster the data belong. Thereby, stability of the partition is usually related to an optimal number of clusters therein. For example, [5] considers stability of a partition with respect to perturbations of the data points, and measures of stability of a cluster are defined as Loevinger's measures. This property of a partition is used to determine a partition with the most appropriate number of clusters. A similar problem is considered in [17]: Does a small change of the sites, e.g., of their position or shape, yield a small change in the corresponding Voronoi cells ? In [8], the Jaccard coefficient, as a similarity measure between sets, is used as the measure of cluster stability, but it is also possible to use some other criteria, like the Rand index, the Hamming distance, the minimal matching distance, and the Variation of Information distance (see, e.g., [14]).

2.1. One-dimensional data points

First, we analyze the separability of clusters in a partition for a one-dimensional data set, since in this case the analysis is simpler, and in addition, a better estimate for the radius of the separability ball can be obtained.

Let $\mathcal{A} = \{a_i \in \mathbb{R} \colon i \in I\}$ be a data set and $z_1 < \cdots < z_k$ the assignment points. Assume that, based on the minimum distance principle, according to Definition 2, a well-separated partition $\Pi = \{\pi(z_1), \ldots, \pi(z_k)\}$ of the set $\mathcal{A}$ is defined by means of the points z_j, $j \in J$, where

$$\pi(z_j) = \{a \in \mathcal{A} \colon |z_j - a| \leq |z_s - a|,\, s \in J\}, \quad j \in J. \tag{10}$$

We should find a separability ball $B(\delta) = \{u \in \mathbb{R} \colon |u| < \delta\}$, such that $\pi(\zeta_j) = \pi(z_j)$ for all $j = 1, \ldots, k$ and all $\zeta_j = z_j + B(\delta)$. The set $z_j + B(\delta)$ is called a separability ball associated with the assignment points z_j.

Let us first give the following auxiliary lemma.

Lemma 1. *Let $z_1, z_2 \in \mathbb{R}$, $z_1 < z_2$ and $\delta > 0$. Then $|p(z_1, z_2) - p(\zeta_1, \zeta_2)| < \delta$ for all $\zeta_1, \zeta_2 \in \mathbb{R}$, such that $\max\{|z_1 - \zeta_1|, |z_2 - \zeta_2|\} < \delta$.*

Proof. The function $p \colon \mathbb{R}^2 \to \mathbb{R}$ defined by $p(x_1, x_2) = \frac{1}{2}(x_1 + x_2)$ satisfies

$$\begin{aligned} |p(z_1, z_2) - p(\zeta_1, \zeta_2)| &= \tfrac{1}{2}|(z_1 - \zeta_1) + (z_2 - \zeta_2)| \\ &\leq \max\{|z_1 - \zeta_1|, |z_2 - \zeta_2|\} < \delta. \end{aligned}$$

□

The following theorem shows how the radius of the separability ball in a well-separated partition can be determined.

Theorem 1. *Let $\mathcal{A} \subset \mathbb{R}$ be a data set and $z_1 < \cdots < z_k$ a set of assignment points which, according to the minimum distance principle, determine a well-separated partition $\Pi = \{\pi(z_1), \ldots, \pi(z_k)\}$ of the set $\mathcal{A}$, and let*

$$0 < \delta = \min_{1 \leq j < s \leq k} \; \min_{a \in \pi(z_j) \cup \pi(z_s)} \left\{ |p(z_j, z_s) - a|, \tfrac{1}{2}(z_s - z_j) \right\}. \tag{11}$$

Then $B(\delta) = \{u \in \mathbb{R} \colon |u| < \delta\}$ is a separability ball of the partition Π and separability balls associated with assignment points are mutually disjoint.

Particularly, if $\zeta_j \in z_j + B(\delta)$, $j \in J$, then the perturbed partition $\hat{\Pi} = \{\pi(\zeta_1), \dots, \pi(\zeta_k)\}$ defined according to the minimum distance principle is well-separated.

Proof. Let $\zeta_j \in z_j + B(\delta)$, $j \in J$. Note first the following equivalences for $a \in \mathcal{A}$

$$a \in \pi(z_1) \quad \Leftrightarrow \quad a < p(z_1, z_2), \tag{12}$$
$$a \in \pi(z_j) \quad \Leftrightarrow \quad p(z_{j-1}, z_j) < a < p(z_j, z_{j+1}),\ j = 2, \dots, k-1, \tag{13}$$
$$a \in \pi(z_k) \quad \Leftrightarrow \quad a > p(z_{k-1}, z_k). \tag{14}$$

Also note that according to (11) and Lemma 1, for $j \in \{1, \dots, k-1\}$ the following holds

$$|p(z_j, z_{j+1}) - a| > \delta, \quad \forall\, a \in \mathcal{A}, \tag{15}$$

$$|p(z_j, z_{j+1}) - p(\zeta_j, \zeta_{j+1})| < \delta, \tag{16}$$

and consequently

$$p(z_j, z_{j+1}) + \delta < a < p(z_j, z_{j+1}) - \delta, \tag{17}$$

$$p(z_j, z_{j+1}) - \delta < p(\zeta_j, \zeta_{j+1}) < p(z_j, z_{j+1}) + \delta. \tag{18}$$

First, let us note that (11) implies $\zeta_i \neq \zeta_j$, $1 \leq i < j \leq k$, and let us show

$$a \neq p(\zeta_j, \zeta_{j+1}) = \tfrac{1}{2}(\zeta_j + \zeta_{j+1}), \text{ for all } a \in \mathcal{A} \text{ and } j = 1, \dots, k-1. \tag{19}$$

Indeed, the existence of a point $a_{i_0} \in \mathcal{A}$, such that $a_{i_0} = p(\zeta_j, \zeta_{j+1})$, for some index $j \in J$, would contradict (15) because of (16).

Next, we show that

$$\pi(z_j) = \pi(\zeta_j), \quad j = 1, \dots, k. \tag{20}$$

First, let us show that $\pi(z_j) \subseteq \pi(\zeta_j)$, $j = 1, \dots, k$. Specially, if $a \in \pi(z_1)$, by using (17) and (18) we obtain

$$a \overset{(17)}{<} p(z_1, z_2) - \delta \overset{(18)}{<} p(\zeta_1, \zeta_2),$$

i.e., $a \in \pi(\zeta_1)$, where

$$\pi(\zeta_1) = \{a \in \mathcal{A} \colon a < p(\zeta_1, \zeta_2)\}.$$

Similarly, one can prove that if $a \in \pi(z_k)$, then $a \in \pi(\zeta_k)$, where

$$\pi(\zeta_k) = \{a \in \mathcal{A} \colon a > p(\zeta_{k-1}, \zeta_k)\}.$$

Let $j \in \{2, \dots, k-1\}$. For $a \in \pi(z_j)$, using (17) and (18) we obtain

$$\begin{aligned} a &\overset{(17)}{>} p(z_{j-1}, z_j) + \delta \overset{(18)}{>} p(\zeta_{j-1}, \zeta_j), \text{ and} \\ a &\overset{(17)}{<} p(z_j, z_{j+1}) - \delta \overset{(18)}{<} p(\zeta_j, \zeta_{j+1}). \end{aligned}$$

Hence, $p(\zeta_{j-1}, \zeta_j) < a < p(\zeta_j, \zeta_{j+1})$, i.e., $a \in \pi(\zeta_j)$, where

$$\pi(\zeta_j) = \{a \in \mathcal{A} \colon p(\zeta_{j-1}, \zeta_j) < a < p(\zeta_j, \zeta_{j+1})\}.$$

Therefore,

$$\pi(z_j) \subseteq \pi(\zeta_j), \quad j \in J. \tag{21}$$

Let us show the opposite inclusion: $\pi(\zeta_j) \subseteq \pi(z_j)$, $j \in J$. Let $j \in J$ be arbitrary and suppose that $a \in \mathcal{A}$ belongs to $\pi(\zeta_j)$. If $a \in \pi(z_j)$, we are done. Suppose $a \in \pi(z_s)$ for some index $s \in J \setminus \{j\}$. Because of (21), a belongs to $\pi(\zeta_s)$. Therefore, $a \in \pi(\zeta_j) \cap \pi(\zeta_s)$. This means that $|a - \zeta_j| = |a - \zeta_s|$, i.e., $a = \frac{1}{2}(\zeta_j + \zeta_s)$. Since $\mathcal{A} \subseteq \mathbb{R}$, numbers s and j are consecutive which contradicts the previously proven claim (19). Thus, claim (20) has also been proved.

Let us now show that the partition $\hat{\Pi} = \{\pi(\zeta_1), \dots, \pi(\zeta_k)\}$ is a well-separated partition according to Definition 2. This is easy to see by using (19) and the implication

$$\emptyset \notin \Pi = \{\pi(z_1), \dots, \pi(z_k)\} \Rightarrow \emptyset \notin \hat{\Pi} = \{\pi(\zeta_1), \dots, \pi(\zeta_k)\},$$

which follows from (21).

Finally, according to Definition 2, $B(\delta)$ is a separability ball of the partition Π, where in accordance with (11), separability balls $z_j + B(\delta)$, $j \in J$, associated with assignment points $z_1, \dots, z_k$, are mutually disjoint. □

Remark 1. Let $\mathcal{A} \subset \mathbb{R}$ be a set of data points and $z_1 < \cdots < z_k$ a set of assignment points. If there exists $1 \le j_0 < s_0 \le k$ and $a_{i_0} \in \pi(z_{j_0}) \cup \pi(z_{s_0})$, such that $a_{i_0} = \frac{1}{2}(z_{j_0} + z_{s_0})$, then a separability ball $B(\delta)$, $\delta > 0$, does not exist, and the partition $\Pi = \{\pi(z_1), \ldots, \pi(z_k)\}$ is not a well-separated partition of the set $\mathcal{A}$.

2.2. n-Dimensional data points

Let a set of data points $\mathcal{A} = \{a_i \in \mathbb{R}^n : i \in I\}$ be given. First, we consider a special case $k = 2$. Assume that for two different assignment points $z_1, z_2 \in \mathbb{R}^n$, based upon the minimum distance principle in accordance with Definition 1, a well-separated partition $\Pi = \{\pi(z_1), \pi(z_2)\}$ of the set $\mathcal{A}$ is defined, where $p = \frac{1}{2}(z_1 + z_2)$.

Proposition 1. *Let $\mathcal{A} \subset \mathbb{R}^n$ be a set of data points and $z_1, z_2 \in \mathbb{R}^n$ two different assignment points, which according to the minimum distance principle define a well-separated partition $\Pi = \{\pi(z_1), \pi(z_2)\}$ of the set $\mathcal{A}$, and let*

$$0 < \delta = \min_{a \in \mathcal{A}} \delta_a, \quad \delta_a = -\mu_1^{(a)} - \mu_2^{(a)} + \sqrt{\left(\mu_1^{(a)} + \mu_2^{(a)}\right)^2 + 2|\phi_a|}, \tag{22}$$

where

$$\phi_a := \langle z_2 - z_1, a - p \rangle,\ \mu_1^{(a)} := \|a - z_1\|,\ \mu_2^{(a)} := \|a - z_2\|,\ p := \tfrac{1}{2}(z_1 + z_2).$$

Then the ball $B(\delta) = \{u \in \mathbb{R}^n : \|u\| < \delta\}$ is a separability ball of the partition Π and separability balls $z_1 + B(\delta)$, $z_2 + B(\delta)$ associated with assignment points z_1, z_2, are disjoint. In particular, if $\zeta_j \in z_j + B(\delta)$, $j = 1, 2$, then the perturbed partition $\hat{\Pi} = \{\pi(\zeta_1), \pi(\zeta_2)\}$, defined according to the minimum distance principle, is well-separated.

Proof. For $u_1, u_2 \in \mathbb{R}^n$, denote $\zeta_j = z_j + u_j$, $j = 1, 2$. Note that in accordance with Definition 1, $\pi(z_1), \pi(z_2) \neq \emptyset$ and

$$a \in \pi(z_1) \Leftrightarrow \phi_a < 0 \quad \text{and} \quad a \in \pi(z_2) \Leftrightarrow \phi_a > 0.$$

First, let us show that if $a \in \pi(z_1)$ ($a \in \pi(z_2)$), then $a \in \pi(\zeta_1)$ ($a \in \pi(\zeta_2)$), for all $\zeta_1 \in z_1 + B(\delta_a)$ ($\zeta_2 \in z_2 + B(\delta_a)$).

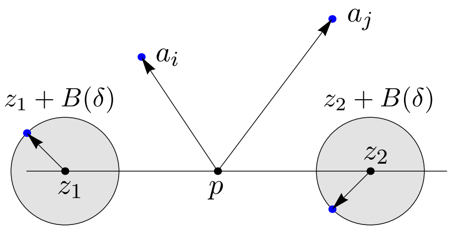


Figure 2: Separability balls associated with assignment points

a) If $a \in \pi(z_1)$, by the Cauchy-Schwartz-Buniakovsky (CSB) inequality, for all $u_j \in B(\delta_a)$, $j = 1, 2$, and $\zeta_j = z_j + u_j \in z_j + B(\delta_a)$, $j = 1, 2$, the following is obtained:

$$\begin{aligned}\langle \zeta_2 - \zeta_1, a - \tfrac{1}{2}(\zeta_1 + \zeta_2)\rangle &= \langle z_2 - z_1 + u_2 - u_1, a - p - \tfrac{1}{2}(u_1 + u_2)\rangle \\ &= -|\phi_a| + \langle u_1, z_1 - a\rangle + \langle u_2, a - z_2\rangle - \tfrac{1}{2}\|u_2\|^2 + \tfrac{1}{2}\|u_1\|^2 \\ &\le -|\phi_a| + \|u_1\|\|z_1 - a\| + \|u_2\|\|z_2 - a\| - \tfrac{1}{2}\|u_2\|^2 + \tfrac{1}{2}\|u_1\|^2 \\ &= -|\phi_a| + \mu_1^{(a)}\|u_1\| + \mu_2^{(a)}\|u_2\| - \tfrac{1}{2}\|u_2\|^2 + \tfrac{1}{2}\|u_1\|^2 \\ &\le -|\phi_a| + \mu_1^{(a)}\|u_1\| + \mu_2^{(a)}\|u_2\| + \tfrac{1}{2}\|u_1\|^2 \\ &< -|\phi_a| + \mu_1^{(a)}\delta_a + \mu_2^{(a)}\delta_a + \tfrac{1}{2}\delta_a^2 \overset{(22)}{=} 0.\end{aligned}$$

Finally,

$$\langle \zeta_2 - \zeta_1, a - \tfrac{1}{2}(\zeta_1 + \zeta_2)\rangle < 0, \tag{23}$$

for all $\zeta_j \in z_j + B(\delta_a)$, $j = 1, 2$. So, if $a \in \pi(z_1)$, then $a \in \pi(\zeta_1)$, for all $\zeta_1 \in z_1 + B(\delta_a)$.

b) If $a \in \pi(z_2)$, by the CSB inequality, for all $u_j \in B(\delta_a)$, $j = 1, 2$, and for $\zeta_j = z_j + u_j \in z_j + B(\delta_a)$, $j = 1, 2$, the following is obtained

$$\begin{aligned}\langle \zeta_2 - \zeta_1, a - \tfrac{1}{2}(\zeta_1 + \zeta_2)\rangle &= \langle z_2 - z_1 + u_2 - u_1, a - p - \tfrac{1}{2}(u_1 + u_2)\rangle \\ &= |\phi_a| - \langle u_1, a - z_1\rangle - \langle u_2, a - z_2\rangle - \tfrac{1}{2}\|u_2\|^2 + \tfrac{1}{2}\|u_1\|^2 \\ &\geq |\phi_a| - \|u_1\|\|z_1 - a\| - \|u_2\|\|z_2 - a\| - \tfrac{1}{2}\|u_2\|^2 + \tfrac{1}{2}\|u_1\|^2 \\ &= |\phi_a| - \mu_1^{(a)}\|u_1\| - \mu_2^{(a)}\|u_2\| - \tfrac{1}{2}\|u_2\|^2 + \tfrac{1}{2}\|u_1\|^2 \\ &\geq |\phi_a| - \mu_1^{(a)}\|u_1\| - \mu_2^{(a)}\|u_2\| - \tfrac{1}{2}\|u_2\|^2 \\ &> |\phi_a| - \mu_1^{(a)}\delta_a - \mu_2^{(a)}\delta_a - \tfrac{1}{2}\delta_a^2 \overset{(22)}{=} 0.\end{aligned}$$

Finally,

$$\langle \zeta_2 - \zeta_1, a - \tfrac{1}{2}(\zeta_1 + \zeta_2)\rangle > 0, \tag{24}$$

for $\zeta_j \in z_j + B(\delta_a)$, $j = 1, 2$. So, if $a \in \pi(z_2)$, then $a \in \pi(\zeta_2)$, for all $\zeta_2 \in z_2 + B(\delta_a)$.

Since

$$\bigcap_{a\in\mathcal{A}} B(\delta_a) = \bigcap_{a\in\mathcal{A}} \{u \in \mathbb{R}^n \colon \|u\| < \delta_a\} = \{u \in \mathbb{R}^n \colon \|u\| < \delta\} = B(\delta),$$

where $\delta = \min_{a\in\mathcal{A}} \delta_a$, for all $\zeta_j \in z_j + B(\delta)$, $j = 1, 2$, we have

$$\langle \zeta_2 - \zeta_1, a - \tfrac{1}{2}(\zeta_1 + \zeta_2)\rangle \neq 0, \forall a \in \mathcal{A}. \tag{25}$$

Let us show that

$$\pi(z_j) = \pi(\zeta_j), \quad j = 1, 2. \tag{26}$$

Let $j \in \{1, 2\}$ and $a \in \pi(z_j)$. Because of $B(\delta) \subseteq B(\delta_a)$, using (23) (resp. (24)), it follows that $a \in \pi(\zeta_j)$, and therefore

$$\pi(z_j) \subseteq \pi(\zeta_j), \quad j = 1, 2. \tag{27}$$

Let us show the opposite inclusion: $\pi(\zeta_j) \subseteq \pi(z_j)$, $j = 1, 2$. Suppose $a \in \pi(\zeta_1)$. If $a \in \pi(z_1)$, we are done. Suppose $a \in \pi(z_2)$. Because of (27), a belongs to $\pi(\zeta_2)$, and therefore, $a \in \pi(\zeta_1) \cap \pi(\zeta_2)$. This means that $\|a - \zeta_1\| = \|a - \zeta_2\|$, i.e., $\langle \zeta_2 - \zeta_1, a - \frac{1}{2}(\zeta_1 + \zeta_2)\rangle = 0$, which contradicts (25).

Analogously, one proves that if $a \in \pi(\zeta_2)$ then $a \in \pi(z_2)$. Thus, claim (26) has also been proved.

In order to prove that the separability balls associated with the assignment points are disjoint, it suffices to show that $\delta < \frac{1}{2}\|z_2 - z_1\|$. Since Π is a well-separated partition, $\phi_a \neq 0$, and since $z_1 \neq z_2$, it follows that $\mu_1^{(a)} + \mu_2^{(a)} \neq 0$. Therefore,

$$-\mu_1^{(a)} - \mu_2^{(a)} + \sqrt{\left(\mu_1^{(a)} + \mu_2^{(a)}\right)^2 + 2|\phi_a|} < \tfrac{|\phi_a|}{\mu_1^{(a)}+\mu_2^{(a)}}. \qquad (28)$$

Namely, by multiplying the inequality $-\mu_1^{(a)} - \mu_2^{(a)} + \sqrt{\left(\mu_1^{(a)} + \mu_2^{(a)}\right)^2 + 2|\phi_a|} > 0$ by $\left(\mu_1^{(a)} + \mu_2^{(a)}\right) > 0$ it follows

$$-\left(\mu_1^{(a)} + \mu_2^{(a)}\right)^2 + \left(\mu_1^{(a)} + \mu_2^{(a)}\right)\sqrt{\left(\mu_1^{(a)} + \mu_2^{(a)}\right)^2 + 2|\phi_a|}$$

which is equivalent to

$$2(\mu_1^{(a)} + \mu_2^{(a)})^2 < (\mu_1^{(a)} + \mu_2^{(a)})^2 + (\mu_1^{(a)} + \mu_2^{(a)})\sqrt{\left(\mu_1^{(a)} + \mu_2^{(a)}\right)^2 + 2|\phi_a|},$$

i.e.,

$$\frac{2(\mu_1^{(a)}+\mu_2^{(a)})^2}{(\mu_1^{(a)}+\mu_2^{(a)})^2+(\mu_1^{(a)}+\mu_2^{(a)})\sqrt{\left(\mu_1^{(a)}+\mu_2^{(a)}\right)^2+2|\phi_a|}} < 1 \Rightarrow \frac{-\mu_1^{(a)}-\mu_2^{(a)}+\sqrt{\left(\mu_1^{(a)}+\mu_2^{(a)}\right)^2+2|\phi_a|}}{\frac{|\phi_a|}{\mu_1^{(a)}+\mu_2^{(a)}}} < 1,$$

from which immediately follows (28).

By the CSB-inequality we obtain

$$\begin{aligned}|\phi_a| &= |\langle z_2 - z_1, a - p\rangle| \leq \|z_2 - z_1\|\|a - p\| \\ &< \tfrac{1}{2}\|z_2 - z_1\|(\|z_1 - a\| + \|z_2 - a\|) = \tfrac{1}{2}\|z_2 - z_1\|(\mu_1^{(a)} + \mu_2^{(a)}),\end{aligned}$$

i.e.,

$$\tfrac{|\phi_a|}{\mu_1^{(a)}+\mu_2^{(a)}} < \tfrac{1}{2}\|z_2 - z_1\|, \; \forall a \in \mathcal{A}.$$

By using (28) we get

$$-\mu_1^{(a)} - \mu_2^{(a)} + \sqrt{\left(\mu_1^{(a)} + \mu_2^{(a)}\right)^2 + 2|\phi_a|} \leq \tfrac{|\phi_a|}{\mu_1^{(a)}+\mu_2^{(a)}} < \tfrac{1}{2}\|z_2 - z_1\|, \; \forall a \in \mathcal{A},$$

i.e.

$$\delta = \min_{a\in\mathcal{A}} \left(-\mu_1^{(a)} - \mu_2^{(a)} + \sqrt{\left(\mu_1^{(a)} + \mu_2^{(a)}\right)^2 + 2|\phi_a|} \right) < \tfrac{1}{2}\|z_2 - z_1\|.$$

Let us also show that $\hat{\Pi} = \{\pi(\zeta_1), \pi(\zeta_2)\}$ is a well-separated partition according to Definition 1. This follows from (25) and the following implication

$$\emptyset \notin \Pi = \{\pi(z_1), \pi(z_2)\} \Rightarrow \emptyset \notin \hat{\Pi} = \{\pi(\zeta_1), \pi(\zeta_2)\},$$

which follows from (27).

□

Remark 2. As mentioned at the beginning of Section 2.1, the estimate of the radius of the separability ball in the one-dimensional case (11) can be obtained much more precisely than in the n-dimensional case (22). Namely, by using (28) and the CSB-inequality we get

$$\begin{aligned} \delta_a &= -\mu_1^{(a)} - \mu_2^{(a)} + \sqrt{\left(\mu_1^{(a)} + \mu_2^{(a)}\right)^2 + 2|\phi_a|} < \tfrac{|\phi_a|}{\mu_1^{(a)}+\mu_2^{(a)}} = \tfrac{|\langle z_2-z_1, a-p\rangle|}{\|z_1-a\|+\|z_2-a\|} \\ &\le \tfrac{\|z_2-z_1\|\|a-p\|}{\|z_1-a\|+\|z_2-a\|} \le \tfrac{(\|z_2-a\|+\|z_1-a\|)\|a-p\|}{\|z_1-a\|+\|z_2-a\|} = \|a-p\|. \end{aligned}$$

Similarly,

$$\begin{aligned} \delta_a &= -\mu_1^{(a)} - \mu_2^{(a)} + \sqrt{\left(\mu_1^{(a)} + \mu_2^{(a)}\right)^2 + 2|\phi_a|} < \tfrac{|\phi_a|}{\mu_1^{(a)}+\mu_2^{(a)}} = \tfrac{|\langle z_2-z_1, a-p\rangle|}{\|z_1-a\|+\|z_2-a\|} \\ &\le \tfrac{1}{2}\tfrac{\|z_2-z_1\|(\|a-z_1\|+\|a-z_2\|)}{\|z_1-a\|+\|z_2-a\|} = \tfrac{1}{2}\|z_2 - z_1\|. \end{aligned}$$

Hence

$$\min_{a\in\mathcal{A}} \delta_a \le \min\left(\min_{a} \|a-p\|, \tfrac{1}{2}\|z_1 - z_2\|\right),$$

and particularly in the one-dimensional case

$$\min_{a\in\mathcal{A}} \delta_a \le \min\left(\min_{a\in\mathcal{A}} |p(z_1,z_2) - a|, \tfrac{1}{2}(z_2 - z_1)\right).$$

The following theorem generalizes Proposition 1 for $m > k \ge 2$ different assignment points.

Theorem 2. *Let $\mathcal{A} \subset \mathbb{R}^n$ be a data set, $z_1, \dots, z_k \in \mathbb{R}^n$ a set of mutually different assignment points which, according to the minimum distance principle, determine a well-separated partition $\Pi = \{\pi(z_1), \dots, \pi(z_k)\}$ of the set $\mathcal{A}$, and let*

$$0 < \delta = \min_{j,s \in \{1,\dots,k\},\, j \neq s} \left\{ \min_{a \in \pi(z_j) \cup \pi(z_s)} \delta_a^{js} \right\},$$

where

$$\delta_a^{js} = -\mu_j^{(a)} - \mu_s^{(a)} + \sqrt{\left(\mu_j^{(a)} + \mu_s^{(a)}\right)^2 + 2|\phi_a|}, \tag{29}$$

and

$$\phi_a := \langle z_s - z_j, a - p \rangle,\ \mu_j^{(a)} := \|a - z_j\|,\ \mu_s^{(a)} := \|a - z_s\|,\ p := \tfrac{1}{2}(z_j + z_s).$$

Then $B(\delta) = \{u \in \mathbb{R} \colon |u| < \delta\}$ is a separability ball of the partition Π and separability balls $z_j + B(\delta)$ associated with assignment points z_j, $j \in J$, are mutually disjoint. In particular, if $\zeta_j \in z_j + B(\delta)$, $j \in J$, then the perturbed partition $\hat{\Pi} = \{\pi(\zeta_1), \dots, \pi(\zeta_k)\}$ defined according to the minimum distance principle is well-separated.

Proof. For each two distinct indices $j, s \in J$, let $\mathcal{A}_{js} := \{a \in \mathcal{A} : a \in \pi(z_j) \cup \pi(z_s)\}$. Consider the partition $\Pi_{js} := \{\pi(z_j), \pi(z_s)\}$ of the set $\mathcal{A}_{js}$ which consists of two clusters. Since Π is well–separated with respect to the data set $\mathcal{A}$, it follows that Π_{js} is well–separated with respect to the data set $\mathcal{A}_{js}$. Let

$$\delta_{js} := \min_{a \in \pi(z_j) \cup \pi(z_s)} \delta_a^{js},$$

where δ_a^{js} is defined by (29). From Proposition 1 it follows that $\delta_{js} > 0$ and the ball $B(\delta_{js})$ is a separability ball of the partition Π_{js}. In addition, separability balls associated with z_j and z_s are disjoint. In particular, for all $\zeta_j \in z_j + B(\delta_{js})$ and all $\zeta_s \in z_s + B(\delta_{js})$ we have

$$\{a \in \mathcal{A}_{js} : \|a - z_j\| < \|a - z_s\|\} = \{a \in \mathcal{A}_{js} : \|a - \zeta_j\| < \|a - \zeta_s\|\}. \tag{30}$$

Note that in accordance with Proposition 1, the perturbed partition $\hat{\Pi}_{js} = \{\pi(\zeta_j), \pi(\zeta_s)\}$ of the set $\mathcal{A}_{js}$ is well–separated.

Let

$$0 < \delta = \min\{\delta_{js} \colon j, s \in J,\, j \neq s\} = \min_{j,s\in J,\, j\neq s} \left\{ \min_{a\in\pi(z_j)\cup\pi(z_s)} \delta_a^{js} \right\}.$$

Choose an arbitrary $j \in J$ and $a \in \pi(z_j)$. One has to show that $a \in \pi(\zeta_j)$ for arbitrary $\zeta_j \in z_j + B(\delta)$.

We want to show that $a \in \pi(\zeta_j)$. Indeed, let $s \neq j$ be an arbitrary index. By the definition of $\mathcal{A}_{js}$ and since $a \in \pi(z_j)$, we have $a \in \mathcal{A}_{js}$ and $\|a - z_j\| \leq \|a - z_s\|$. But the partition Π is well-separated, hence the equality in the previous inequality cannot hold, i.e., $\|a - z_j\| < \|a - z_s\|$. By (30), this implies that $\|a - \zeta_j\| < \|a - \zeta_s\|$. Since $s \neq j$ was an arbitrary index, we have

$$a \in \{x \in \mathcal{A} : \|x - \zeta_j\| < \|x - \zeta_s\| \, \forall s \neq j\} = \pi(\zeta_j).$$

Thus $\pi(z_j) \subseteq \pi(\zeta_j)$. This is true for all indices $j \in J$ since j was arbitrary.

It remains to show that $\pi(\zeta_j) \subseteq \pi(z_j)$ for all $j \in J$. Inclusions $\pi(z_j) \subseteq \pi(\zeta_j)$ for all $j \in J$ imply the inclusion $\mathcal{A} = \bigcup_{j=1}^{k} \pi(z_j) \subseteq \bigcup_{j=1}^{k} \pi(\zeta_j)$. Therefore, given an index $j \in J$ and a data point $a \in \pi(\zeta_j)$, since we have $a \in \mathcal{A}$, it follows that $a \in \pi(z_s)$ for some index s. If $s = j$, then $a \in \pi(z_j)$, proving the statement. Otherwise, $a \in \pi(z_s)$ and by the inclusion $\pi(z_s) \subseteq \pi(\zeta_s)$ it follows that $a \in \pi(\zeta_s)$. Thus $a \in \pi(\zeta_j) \cap \pi(\zeta_s)$, where $j \neq s$. This implies that $\|a - \zeta_j\| = \|a - \zeta_s\|$, contradicting the previously proven statement that the perturbed partition $\hat{\Pi}_{js}$ is well-separated. □

Remark 3. The previous analysis was conducted for the case of spherical clustering by using the squared Euclidean distance. Similarly, the clustering problem could be considered by using the Mahalanobis distance-like function [6], but using other distance-like functions that are not generated by some scalar product would require entirely different techniques. Let us mention that some of these techniques may be in the spirit of [17, 18].

3. A possible application

Knowing the separability ball for some partition of the set $\mathcal{A}$ gives an insight into the internal structure of the partition and the measure of separability and compactness of clusters therein. Theoretical properties of cluster separability open up more possibilities for application in cluster analysis. One possibility may be to try to create a new validity index for searching for a partition with the most appropriate number of clusters. Let us mention that a candidate for such an index has already been developed. Roughly speaking, it is a slight variation of the separability radius, multiplied by a certain scaling factor which takes into account the objective function and the number of data points in slightly modified clusters. Experiments show that this index has promising potential.

Furthermore, it was shown that the radius of the separability ball is in correlation with the objective function value while applying the $k-$means algorithm. In order to illustrate that, in this section we consider the behavior of the radius of the cluster separability ball while applying the $k-$means algorithm.

Let $\mathcal{A} \subset \mathbb{R}^n$ be the set which should be partitioned into $1 \leq k \leq m$ nonempty disjoint clusters by using the squared Euclidean distance. The k-means algorithm (see, e.g., [11, 28, 29]) is the most popular algorithm for searching for a locally optimal partition and it can be described by two steps which are iteratively repeated.

Step 1 For each set of mutually different assignment points $z_1, \ldots, z_k \in \mathbb{R}^n$, the set $\mathcal{A}$ should be divided into k disjoint clusters $\pi_1, \ldots, \pi_k$ by using the minimum distance principle

$$\pi_j = \{a \in \mathcal{A} : \|z_j - a\| \leq \|z_s - a\|, \forall s \in J\}, \qquad j \in J. \tag{31}$$

Step 2 Given a partition $\Pi = \{\pi_1, \ldots, \pi_k\}$ of the set $\mathcal{A}$, one can define the corresponding centroids by

$$c_j = \underset{x \in \operatorname{conv} \pi_j}{\operatorname{argmin}} \sum_{a_i \in \pi_j} \|x - a_i\|^2 = \tfrac{1}{|\pi_j|} \sum_{a_i \in \pi_j} a_i, \quad j = 1, \ldots, k. \tag{32}$$

By using a good initial approximation, this method can provide an acceptable solution [26, 30]. In each step of the k-means algorithm, the value of the objective function does not increase. One of the problems with the $k-$means algorithm is that empty clusters can be obtained if no points are allocated to a cluster during the assignment step. In such situation re-running the algorithm with a new initial approximation is usually recommended.

In case of the squared Euclidean distance the dual objective function

$$\mathcal{G}(\Pi) = \sum_{j=1}^{k} |\pi_j| \, \|c - c_j\|^2, \tag{33}$$

can also be considered [7, 25], where c_j are centroids of clusters and c is the centroid of the whole set of data points $\mathcal{A}$, for which $c = \sum_{j=1}^{k} \frac{|\pi_j|}{m} c_j$, holds. One can show that [7, 25]

$$\underset{\Pi \in \mathcal{P}(\mathcal{A},k)}{\operatorname{argmin}} \mathcal{F}(\Pi) = \underset{\Pi \in \mathcal{P}(\mathcal{A},k)}{\operatorname{argmax}} \mathcal{G}(\Pi),$$

and that in each step of the k-means algorithm the value of the dual objective function does not decrease.

Since on the interval $[0, \infty\rangle$ the function $f_\alpha(x) = -x + \sqrt{\alpha + x^2}$, $\alpha > 0$, decreases, $f_\alpha(x) \le f_\alpha(0) = \sqrt{\alpha}$, and $\lim\limits_{x \to +\infty} f_\alpha(x) = 0$, the radius δ of the separability ball can be estimated as

$$\delta = \min_{\substack{j,s \in \{1,\dots,k\} \\ j \neq s}} \left\{ \min_{a \in \pi(z_j) \cup \pi(z_s)} \delta_a^{js} \right\} \le \min_{\substack{j,s \in \{1,\dots,k\} \\ j \neq s}} \sqrt{2 |\langle c_s - c_j, a - p \rangle|},$$

where

$$\delta_a^{js} = -\mu_j^{(a)} - \mu_s^{(a)} + \sqrt{\left(\mu_j^{(a)} + \mu_s^{(a)}\right)^2 + 2|\phi_a|},$$

and

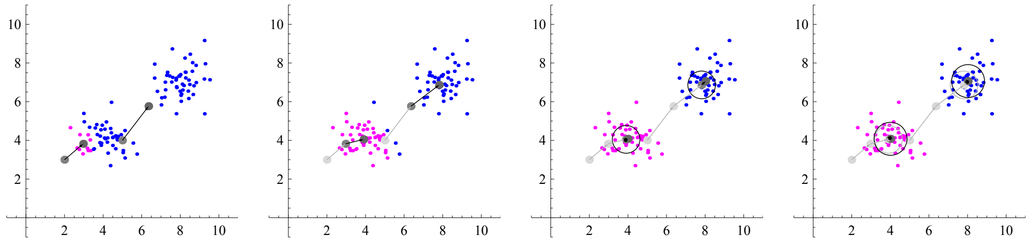
$$\phi_a = \langle z_s - z_j, a - p \rangle, \ \mu_j^{(a)} = \|a - z_j\|, \ \mu_s^{(a)} = \|a - z_s\|, \ p = \tfrac{1}{2}(z_j + z_s).$$

Note that for $k = 2$ there holds

$$\delta \le \sqrt{2 \|c_2 - c_1\|} \min_{a \in \mathcal{A}} \kappa_a, \qquad \kappa_a = \frac{|\langle c_2 - c_1, a - p \rangle|}{\|c_2 - c_1\|,} \tag{34}$$

and it can be associated with the dual objective function $\mathcal{G}$ (see Example 2).

In the following simple example, we consider partitioning of the set $\mathcal{A} \subset \mathbb{R}^2$ into two clusters π_1, π_2, and analyze the behavior of separability balls during the k-means algorithm.



(a) Iteration 1 $\delta = 10^{-6}\,(d = 10.0)$ (b) Iteration 2 $\delta = .02\,(d = 15.2)$ (c) Iteration 3 $\delta = .72\,(d = 23.2)$ (d) Iteration 4 $\delta = .85\,(d = 25.0)$

Figure 3: The movement of the distance $d = \|c_2 - c_1\|^2$ between the centroids and the radius of the separability ball δ in each iteration of the k-means algorithm

Example 2. *Two points $C_1 = (4, 4)$, $C_2 = (8, 7)$ were chosen in the square $[0, 10]^2$, and in the neighborhood of each point 50 random points were generated by using Gaussian distributions. In this way, we obtained the original partition $\Pi = \{\pi_1, \pi_2\}$ and the set of data points $\mathcal{A} = \pi_1 \cup \pi_2$ with $m = 100$ data points.*

The k-means algorithm starts with two different assignment points $z_1 = (2, 3)$, $z_2 = (5, 4)$, and by using the minimum distance principle the clusters $\pi_1(c_1), \pi_2(c_2)$ with centroids c_1, c_2 are obtained. In this case, the dual objective function is

$$\begin{aligned}\mathcal{G}(\pi_1, \pi_2) &= |\pi_1|\|c - c_1\|^2 + |\pi_2|\|c - c_2\|^2 \\ &= |\pi_1| \left\| \tfrac{|\pi_2|}{m} c_2 - \tfrac{|\pi_2|}{m} c_1 \right\|^2 + |\pi_2| \left\| \tfrac{|\pi_1|}{m} c_1 - \tfrac{|\pi_1|}{m} c_2 \right\|^2 \\ &= \tfrac{|\pi_1||\pi_2|}{m} \|c_2 - c_1\|^2 = \tfrac{1}{2} H(|\pi_1|, |\pi_2|) \|c_2 - c_1\|^2, \end{aligned} \tag{35}$$

where $m = |\pi_1| + |\pi_2|$ and $c = \frac{|\pi_1|}{m} c_1 + \frac{|\pi_2|}{m} c_2$, is the centroid of the whole set $\mathcal{A}$ (see [11]). $H(|\pi_1|, |\pi_2|)$ is the harmonic mean of the numbers of data points in clusters π_1 and π_2. Using (35) in (34), we obtain

$$\delta \leq \sqrt{2\sqrt{2}} \sqrt[4]{\tfrac{\mathcal{G}(\pi_1, \pi_2)}{H(|\pi_1|, |\pi_2|)}} \min_{a \in \mathcal{A}} \kappa_a. \tag{36}$$

Note that formula (36) describes the connection between the radius of the separability ball δ and the value of the dual objective function. As can be seen in Fig. 3, the distance $d = \|c_2 - c_1\|^2$ between the centroids, and the radius of the separability ball δ at the end of the k-means algorithm increase. Namely, then the value of the dual objective function increases, too.

4. Conclusions

It can be expected that the assessment of the separability ball size of the partition can be a very useful tool in cluster analysis. Knowing the separability ball for some partition of the set $\mathcal{A}$ gives us an insight into the internal structure of the partition and the measure of separability and compactness of clusters therein (how well separated and how homogeneous the clusters are).

Further research could be directed toward to the applications of cluster separability. For example, construction of a new validity index for searching for a partition with the most appropriate number of clusters can be considered. For more details about this, see the beginning of Section 3.

Acknowledgements.

The authors would like to thank Professor Šime Ungar (Department of Mathematics, University of Osijek) for his useful comments and remarks. We are also thankful to anonymous referees and journal editors for their careful reading of the paper and insightful comments that helped us improve the paper. This work is supported by the Ministry of Science, Education and Sports, Republic of Croatia.

[1] Aurenhammer, F., Klein, R., 2000. Voronoi diagrams, in: Sack, J., Urrutia, G. (Eds.), Handbook of Computational Geometry, Chapter V. Elsevier Science Publishing, 201–290.

[2] Bandyopadhyay, S., Saha, S., 2013. Unsupervised Classification: Similarity Measures, Classical and Metaheuristic Approaches, and Applications. Springer, Berlin Heidelberg, Germany.

[3] Bezdek, J.C., Keller, J., Krisnapuram, R., Pal, N.R., Dubois, D., Prade, H., 2005. Fuzzy models and algorithms for pattern recognition and image processing. Springer, New York, USA.

[4] Ben-Hur, A., Elisseeff, A., Guyon, I., 2002. A stability based method for discovering structure in clustered data, in: Pacific Symposium on Biocomputing, Lihue, Hawaii, USA. 6–17.

[5] Bertrand, P., Bel Mufti, G., 2006. Loevingers measures of rule quality for assessing cluster stability. Computational Statistics and Data Analysis 50, 992–1015.

[6] Durak, B., 2011. A Classification Algorithm Using Mahalanobis Distances Clustering of Data with Applications on Biomedical Data Set. Ph.D. thesis. The Graduate School of Natural and Applied Sciences of Middle East Technical University.

[7] Gan, G., Ma, C., Wu, J., 2007. Data Clustering: Theory, Algorithms, and Applications. SIAM, Philadelphia, USA.

[8] Hennig, C., 2007. Cluster-wise assessment of cluster stability. Computational Statistics & Data Analysis 52, 258–271.

[9] Iyigun, C., Ben-Israel, A., 2010. A generalized Weiszfeld method for the multi-facility location problem. Operations Research Letters 38, 207–214.

[10] Kim, D., Lee, K.H., Leeb, D., 2004. On cluster validity index for estimation of the optimal number of fuzzy clusters. Pattern Recognition 37, 2009–2025.

[11] Kogan, J., 2007. Introduction to Clustering Large and High-dimensional Data. Cambridge University Press, New York, USA.

[12] Lange, T., Roth, V., Braun, M., Buhmann, J., 2004. Stability-based validation of clustering solutions. Neural Computation 16, 1299–1323.

[13] Leisch, F., 2006. A toolbox for k-centroids cluster analysis. Computational Statistics & Data Analysis 51, 526–544.

[14] von Luxburg, U., 2009. Clustering stability: an overview. Foundations and Trends in Machine Learning 2, 235–274.

[15] Okabe, A., Boots, B., Sugihara, K., Chiu, S., N., 2000. Spatial Tessellations: Concepts and Applications of Voronoi diagrams. John Wiley & Sons, Chichester, UK.

[16] Pascual, D., Pla, F., Sánchez, J.S., 2010. Cluster stability assessment based on theoretic information measures. Pattern Recognition Letters 31, 454–461.

[17] Reem, D., 2011. The geometric stability of Voronoi diagrams with respect to small changes of the sites, in: Proceedings of the 27th Annual ACM Symposium on Computational Geometry (SoCG 2011), 254–263.

[18] Reem, D., 2012. The geometric stability of Voronoi diagrams in normed spaces which are not uniformly convex, in: arXiv:1212.1094 [cs.CG].

[19] Sabo, K., 2014. Center-based l_1-clustering method. International Journal of Applied Mathematics and Computer Science 24, 151–163.

[20] Sabo, K., Scitovski, R., Vazler, I., 2013. One-dimensional center-based l_1-clustering method. Optimization Letters 7, 5–22.

[21] Sabo, K., Scitovski, R., Taler, P., 2012. Uniform distribution of the number of voters per constituency on the basis of a mathematical model (in Croatian). Hrvatska i komparativna javna uprava 14, 229–249.

[22] Scitovski, R., Sabo, K., 2014. Analysis of the k-means algorithm in the case of data points occurring on the border of two or more clusters. Knowledge-Based Systems 57, 1–7.

[23] Scitovski, R., Scitovski, S., 2013. A fast partitioning algorithm and its application to earthquake investigation. Computers and Geosciences 59, 124–131.

[24] Shamir, O., Tishby, N., 2010. Model selection and stability in k-means clustering. Machine Learning 80, 213–243.

[25] Späth, H., 1983. Cluster-Formation und Analyse. R. Oldenburg Verlag, München.

[26] Steinley, D., Brusco, M.J., 2007. Initializing k-means batch clustering: a critical evaluation of several techniques. Journal of Classification 24, 99–121.

[27] Su, Z., Kogan, J., Nicholas, C., 2010. Constrained clustering with k-means type algorithms, in: Berry, M.W., Kogan, J. (Eds.), Text Mining: Applications and Theory. Wiley, Chichester, 81–103.

[28] Teboulle, M., 2007. A unified continuous optimization framework for center-based clustering methods. Journal of Machine Learning Research 8, 65–102.

[29] Theodoridis, S., Koutroumbas, K.,2009. Pattern Recognition. Academic Press, Burlington.

[30] Volkovich, V., Kogan, J., Nicholas, C., 2007. Building initial partitions through sampling techniques. European Journal of Operational Research 183, 1097–1105.

[31] Wang, W., Zhang, Y., 2007. On fuzzy cluster validity indices. Fuzzy Sets and Systems 158, 2095–2117.

[32] Xie, X.L., Beni, G., 1991. A validity measure for fuzzy clustering. IEEE Trans. Pattern Anal. Mach. Intell. 13(8), 841–847.