Robust k-means method based on minimizing differentiable estimates of mean, insensitive to outliers

Zaur Shibzukhov
Institute Mathematics and Informatics
Moscow Pedagogical State Univercity and VNII GO CHS
Moscow, Russia
szport@gmail.com

Mukhamed Kazakov
Institute of Applied Mathematics and Automation
Kabardino-Balkarian Scientific Center RAS
Nalchik, Russia
kasakow.muchamed@gmail.com

Dmitriy Dimitrichenko
Institute of Applied Mathematics and Automation
Kabardino-Balkarian Scientific Center RAS
Nalchik, Russia
dimdp@rambler.ru

***Abstract*—The article presents a new approach to the problem of searching for cluster centers, based on minimizing differentiable estimates of the average value, insensitive to outliers. It is at the level of the initial mathematical formulation of the problem to lay the stability of the solution with respect to outliers in the data. The search for cluster centers is carried out using the Mahalanobis distance. The proposed algorithm is based on an iterative reweighting scheme. At each step, the problem of searching for cluster centers based on an algorithm with weights of examples is solved. The weights of the examples correspond to the values of the partial derivatives of a function that estimates the average value and is insensitive to outliers. The weights obtained in this way suppress the effect of emissions.**

Keywords—robust clustering, robust mean estimate, iteratively reweighted algorithm

I. Introduction

The problem of searching for cluster centers has been in the field of attention of researchers for many years [1], [2], [3].

The classical method for searching for centers and covariance matrices of clusters can be based on solving the following minimization problem:

$$\mathbf{c}_1^*, \dots, \mathbf{c}_K^* = \arg\min_{\mathbf{c}_1,\dots,\mathbf{c}_K} \frac{1}{N}\sum_{k=1}^{N} \min_{j=1,\dots,K} d(\mathbf{x}_k; \mathbf{c}_j, \mathbf{S}_j) \qquad (1)$$

where $\mathbf{c}_1, \dots, \mathbf{c}_K$ are cluster centers, $\mathbf{S}_1, \dots, \mathbf{S}_K$ are covariance matrices,

$$d(\mathbf{x}; \mathbf{c}, \mathbf{S}) = \ln|\mathbf{S}| + (\mathbf{x}-\mathbf{c})'\mathbf{S}^{-1}(\mathbf{x}-\mathbf{c})$$

is the square of the Mahalanobis distance with the covariance matrix $\mathbf{S}$ between the points $\mathbf{x}$ and $\mathbf{c}$.

This statement of the problem is based on the assumption that the points of the j-th cluster obey a multidimensional normal distribution with a density

$$p(\mathbf{x}; \mathbf{c}, \mathbf{S}) \propto \frac{1}{\sqrt{|\mathbf{S}|}} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{c})'\mathbf{S}^{-1}(\mathbf{x}-\mathbf{c})},$$

an arbitrary point $\mathbf{x}$ refers to the cluster with the number

$$j(\mathbf{x}) = \arg\max_{j=1,\dots,K} p(\mathbf{x}; \mathbf{c}_j, \mathbf{S}_j).$$

The problem (1) is reduced to solving systems of equations

$$\begin{cases} \mathbf{c}_j = \dfrac{1}{|\mathbf{I}_j|}\displaystyle\sum_{k\in\mathbf{I}_j} \mathbf{x}_k \\ \mathbf{S}_j = \dfrac{1}{|\mathbf{I}_j|}\displaystyle\sum_{k\in\mathbf{I}_j} (\mathbf{x}_k - \mathbf{c}_j)'(\mathbf{x}_k - \mathbf{c}_j), \end{cases} \qquad (2)$$

where $\mathbf{I}_j \subset \{1, \dots, N\}$ are indices of points falling into the j-th cluster.

The following iterative procedure underlies the extended k-means algorithm:

$$\begin{cases} \mathbf{c}_{j,t+1} = \dfrac{1}{|\mathbf{I}_{j,t}|}\displaystyle\sum_{k\in\mathbf{I}_{j,t}} \mathbf{x}_k \\ \mathbf{S}_{j,t+1} = \dfrac{1}{|\mathbf{I}_{j,t}|}\displaystyle\sum_{k\in\mathbf{I}_{j,t}} (\mathbf{x}_k - \mathbf{c}_{j,t})'(\mathbf{x}_k - \mathbf{c}_{j,t}), \end{cases} \qquad (3)$$

where $\mathbf{I}_{j,t}$ are indices of points falling into the j-th cluster at the t-th step. Initial values of cluster centers $\mathbf{c}_{1,0}, \dots, \mathbf{c}_{K,0}$ and covariance matrices $\mathbf{S}_{1,0}, \dots, \mathbf{S}_{K,0}$ are set before the iteration procedure (3).

A significant distortion of the results of the algorithm may appear if the empirical distribution $\{D(\mathbf{x}_1), \dots, D(\mathbf{x}_N)\}$, where

$$D(\mathbf{x}) = D(\mathbf{x}; \mathbf{c}_1, \dots, \mathbf{c}_K; \mathbf{S}_1, \dots, \mathbf{S}_K) = \min_{j=1,\dots,K} d(\mathbf{x}; \mathbf{c}_j, \mathbf{S}_j)$$

contains oitliers.

II. The classic method of overcoming the effects of outliers

The classical method for solving the problem of emissions is based on the replacement of the function $d(\mathbf{x}; \mathbf{c}, \mathbf{S})$ with

$$d_\varrho(\mathbf{x}; \mathbf{c}, \mathbf{S}) = \ln|\mathbf{S}| + \varrho((\mathbf{x}-\mathbf{c})'\mathbf{S}^{-1}(\mathbf{x}-\mathbf{c})),$$

where $\varrho(r)$ is a function to suppress the effects of outliers. It corresponds to the probability distribution of points with density

$$p(\mathbf{x}; \mathbf{c}, \mathbf{S}) \propto \frac{1}{\sqrt{|\mathbf{S}|}} e^{-\frac{1}{2}\varrho((\mathbf{x}-\mathbf{c})'\mathbf{S}^{-1}(\mathbf{x}-\mathbf{c}))}$$

The optimization task has the form:

$$\mathbf{c}_1^*, \dots, \mathbf{c}_K^* = \arg\min_{\mathbf{c}_1,\dots,\mathbf{c}_K} \frac{1}{N}\sum_{k=1}^{N} D_\varrho(\mathbf{x}_k), \qquad (4)$$

where

$$D_\varrho(\mathbf{x}) = D_\varrho(\mathbf{x}; \mathbf{c}_1, \dots, \mathbf{c}_K; \mathbf{S}_1, \dots, \mathbf{S}_K) = \min_{j=1,\dots,K} d_\varrho(\mathbf{x}; \mathbf{c}_j, \mathbf{S}_j)$$

The ϱ function is introduced in order to achieve a relative decrease in the large values of the square of the Mahalanobis function. An example is the function $\varrho(r) = H(\sqrt{r})$, where H is the Huber function:

$$H(r) = \begin{cases} \frac{1}{2}r^2, & \text{if } r \le c \\ rc - \frac{1}{2}c^2, & \text{if } r < c. \end{cases}$$

Along with the Huber function, you can also use the function $S(r) = \sqrt{c^2 + r^2} - c$, which, unlike it, has a continuous 2nd-order derivative.

The problem (4) can be reduced to solving a system of equations:

$$\begin{cases} \mathbf{c}_j = \frac{1}{V_j}\sum_{k\in\mathbf{I}_j} v_k \mathbf{x}_k, \qquad V_j = \sum_{k\in\mathbf{I}_j} v_k \\ \mathbf{S}_j = \frac{1}{V_j}\sum_{k\in\mathbf{I}_j} v_k (\mathbf{x}_k - \mathbf{c}_j)'(\mathbf{x}_k - \mathbf{c}_j), \end{cases} \tag{5}$$

where $v_k = \psi(D_\varrho(\mathbf{x}_k))$, $\psi(r) = \varrho'(r)$.

For the solution to be unique, it is necessary that $\varrho'(r)$ be non-decreasing. But it follows from this that it is enough to make outliers of the order $\frac{1}{n+1}$-th part of the set of points in order to break the robustness of such a method [4]. Nevertheless, if the matrices $\mathbf{S}_1, \dots, \mathbf{S}_K$ are given, then the problem of finding the centers $\mathbf{c}_1, \dots, \mathbf{c}_K$ is robust. The loss of robustness is precisely connected with the evaluation of the matrices $\mathbf{S}_1, \dots, \mathbf{S}_K$.

A fairly comprehensive overview of other methods can be found in [5], [6].

III. The Principle of Minimizing Differentiable Averages, Insensitive to Outliers

In this paper, we propose a new approach based on replacing the arithmetic mean in (1) with a robust differentiable mean estimate of $\mathsf{M}\{z_1, \dots, z_N\}$, which will be insensitive to outliers. Such a replacement will allow, at the level of the mathematical formulation of the problem, to lay the foundation for the stability of the solution of the problem. This is precisely the novelty of the proposed approach. Since the empirical distribution of the squares of the distances of the Mahalanobis from the points to the center of the nearest cluster may contain outliers, so the arithmetic mean value turns out to be distorted. As a consequence of this, the positions of the centers of the clusters may be displaced. The use of an outliers-insensitive average estimate can avoid distortion.

The differentiability of the estimate of the average value, insensitive to outliers, allows the use of gradient minimization algorithms to search for cluster centers.

Thus, in terms of outliers, it is proposed to search for $\mathbf{c}_1^*, \dots, \mathbf{c}_K^*$ and $\mathbf{S}_1^*, \dots, \mathbf{S}_K^*$, minimizing the functional

$$Q(\mathbf{c}_1, \dots, \mathbf{c}_K; \mathbf{S}_1, \dots, \mathbf{S}_K) = \mathsf{M}\{D_1(\mathbf{c}_1, \dots, \mathbf{c}_K; \mathbf{S}_1, \dots, \mathbf{S}_K), \dots, D_N(\mathbf{c}_1, \dots, \mathbf{c}_K; \mathbf{S}_1, \dots, \mathbf{S}_K)\},$$

where

$$D_k(\mathbf{c}_1, \dots, \mathbf{c}_K; \mathbf{S}_1, \dots, \mathbf{S}_K) = D(\mathbf{x}_k; \mathbf{c}_1, \dots, \mathbf{c}_K; \mathbf{S}_1, \dots, \mathbf{S}_K).$$

Due to differentiability of $\mathsf{M}\{z_1, \dots, z_N\}$ the desired centers $\mathbf{c}_1^*, \dots, \mathbf{c}_K^*$ and the matrices $\mathbf{S}_1^*, \dots, \mathbf{S}_K^*$ are the solutions of the system of nonlinear equations:

$$\begin{cases} z_k = D_k(\mathbf{c}_1, \dots, \mathbf{c}_K; \mathbf{S}_1, \dots, \mathbf{S}_K), & k = 1, \dots, N \\ \mathbf{v} = \nabla\mathsf{M}\{z_1, \dots, z_N\} & \\ \mathbf{c}_j = \frac{1}{V_j}\sum_{k\in\mathbf{I}_j} v_k \mathbf{x}_k, & j = 1, \dots, K \\ \mathbf{S}_j = \frac{1}{V_j}\sum_{k\in\mathbf{I}_j} v_k (\mathbf{x}_k - \mathbf{c}_j)'(\mathbf{x}_k - \mathbf{c}_j), & j = 1, \dots, K \end{cases} \tag{6}$$

The vector of sample weights $\mathbf{v}$ for $\mathbf{c}_j = \mathbf{c}_j^*$ and $\mathbf{S}_j = \mathbf{S}_j^*$ can also be used as an estimate of the significance of points. Since $v_1 + \dots + v_N = 1$, the outliers will correspond to the points with the lowest values of the weights.

Stability with respect to outliers is achieved due to the fact that the weights of the points corresponding to outliers are significantly less than the weights of the points that are not outliers. It is also important that the point weight decreases as the absolute value of the difference between $\bar{z} = \nabla\mathsf{M}\{z_1, \dots, z_N\}$ and z_k increases. Such properties are a natural consequence of the robustness of mean estimates.

IV. Outliers Insensitive Average Estimates

Such estimates can be constructed in at least two ways.

The first method is based on the approximation of the median based on the M-mean [7], [8]:

$$\mathsf{M}_\rho\{z_1, \dots, z_N\} = \arg\min_u \sum_{k=1}^{N} \rho(z_k - u),$$

where ρ is twice differentiable strictly convex function with a minimum at zero. The M-mean defined in this way has partial derivatives:

$$\frac{\partial \mathsf{M}_\rho}{\partial z_k} = \frac{\rho''(z_k - \bar{z}_\rho)}{\rho''(z_1 - \bar{z}_\rho) + \dots + \rho''(z_N - \bar{z}_\rho)},$$

where $\bar{z}_\rho = \mathsf{M}_\rho\{z_1, \dots, z_N\}$.

For example, if you take the function $\rho(r) = \sqrt{\varepsilon^2 + r^2} - \varepsilon$, then for sufficiently small values $\varepsilon > 0$, you can get an approximate and smoothed version of the median. Choosing a sufficiently small value of ε, we can ensure that the value $\partial\mathsf{M}_\rho/\partial z_k$ is negligible for those values z_k that are far from the average value $\bar{z}_\rho$.

Smoothed variant of α-quantile can be built based on the function

$$\rho_\alpha(r) = \begin{cases} \alpha\rho(r), & \text{if } r > 0 \\ \frac{1}{2}\big(\alpha\rho(0_+) + (1-\alpha)\rho(0_+)\big), & \text{if } r = 0 \\ (1-\alpha)\rho(r), & \text{if } r < 0, \end{cases} \tag{7}$$

where $\rho(r)$ is a function for smoothed variant of median.

The second method is based on the use of a censored arithmetic mean, in which the threshold value is estimated using a smoothed version of the α-quantile:

$$\mathrm{WM}_{\rho,\alpha}\{z_1,\dots,z_N\}=\frac{1}{N}\sum_{k=1}^{N}\min\{z_k,\bar{z}_{\rho_\alpha}\}. \qquad (8)$$

Partial derivatives are of the form:

$$\frac{\partial \mathrm{WM}_{\rho_\alpha}}{\partial z_k}=\begin{cases}\frac{1}{N}+\frac{m}{N}\frac{\partial \mathsf{M}_{\rho_\alpha}}{\partial z_k}, & \text{if } z_k<\bar{z}_{\rho_\alpha}\\ \frac{m}{N}\frac{\partial \mathsf{M}_{\rho_\alpha}}{\partial z_k}, & \text{if } z_k\geq\bar{z}_{\rho_\alpha},\end{cases}$$

where m is equal to the number of $z_k\geq\bar{z}_{\rho_\alpha}$. In both cases $\frac{\partial \mathsf{M}}{\partial z_k}\geq 0$ and

$$\frac{\partial \mathsf{M}}{\partial z_1}+\dots+\frac{\partial \mathsf{M}}{\partial z_N}=1.$$

The third method takes a different approach to censoring values. Let's define a truncated version of a quadratic function:

$$r_c^2=\begin{cases}r^2, & \text{if } |r|\leq c\\ c^2, & \text{if } |r|>c.\end{cases}$$

With its help we define

$$\tilde{z}_\alpha=\mathsf{TM}_{\rho_\alpha}\{z_1,\dots,z_N\}=$$

$$=\arg\min_u\left\{\sum_{|z_k-u|\leq\bar{c}_\alpha}(z_k-u)^2+\sum_{|z_k-u|>\bar{c}_\alpha}\bar{c}_\alpha^2\right\},$$

where $\bar{c}_\alpha^2=\mathsf{M}_{\rho_\alpha}\{v_1,\dots,v_N\}$, $v_k=(z_k-u)^2$.

From the definition we get a recurrence relation for calculating $\tilde{z}_\alpha$:

$$u_{t+1}=\sum_{|z_k-u_t|\leq\bar{c}_{\alpha,t}}\left(\frac{1}{N}+\frac{m}{N}\frac{\partial \mathsf{M}_{\rho_\alpha}}{\partial v_{k,t}}\right)z_k+\sum_{|z_k-u_t|>\bar{c}_{\alpha,t}}\frac{m}{N}\frac{\partial \mathsf{M}_{\rho_\alpha}}{\partial v_{k,t}}z_k,$$

where $\bar{c}_{\alpha,t}^2=\mathsf{M}_{\rho_\alpha}\{v_{1,t},\dots,v_{N,t}\}$, $v_{k,t}=(z_k-u_t)^2$, m is the number of values z_k: $|z_k-u_t|>\bar{c}_{\alpha,t}$.

In the limit, we obtain the weighted arithmetic mean:

$$\tilde{z}_\alpha=\sum_{k=1}^{N}v_k\,z_k,$$

where

$$v_k=\begin{cases}\frac{1}{N}+\frac{m}{N}\frac{\partial \mathsf{M}_{\rho_\alpha}}{\partial v_k}, & \text{if } |z_k-\tilde{z}_\alpha|\leq\bar{c}_\alpha\\ \frac{m}{N}\frac{\partial \mathsf{M}_{\rho_\alpha}}{\partial v_k}, & \text{if } |z_k-\tilde{z}_\alpha|>\bar{c}_\alpha,\end{cases}$$

and $\bar{c}_\alpha^2=\mathsf{M}_{\rho_\alpha}\{v_1,\dots,v_N\}$, $v_k=(z_k-\tilde{z}_\alpha)^2$, m is the number of values z_k: $|z_k-\tilde{z}_\alpha|>\bar{c}_\alpha$. Note that

$$\sum_{k=1}^{N}v_k=1.$$

In this situation the system of equations (6) should be rewrited as follow

$$\begin{cases}z_k=D_k(\mathbf{c}_1,\dots,\mathbf{c}_K;\mathbf{S}_1,\dots,\mathbf{S}_K), & k=1,\dots,N\\ \mathbf{c}_j=\frac{1}{V_j}\sum_{k\in\mathbf{I}_j}v_k\mathbf{x}_k, & j=1,\dots,K\\ \mathbf{S}_j=\frac{1}{V_j}\sum_{k\in\mathbf{I}_j}v_k\,(\mathbf{x}_k-\mathbf{c}_j)'(\mathbf{x}_k-\mathbf{c}_j), & j=1,\dots,K,\end{cases}$$

where

$$G_j=\sum_{k\in\mathbf{I}_j}\gamma_k.$$

V. The Algorithm

To search $\mathbf{c}_1^*,\dots,\mathbf{c}_K^*$ and $\mathbf{S}_1^*,\dots,\mathbf{S}_K^*$ we apply an iterative scheme that corresponds to the analog of the Jacobi method for solving the system of nonlinear equations (6).

The initial positions of the centers are selected in some way, for example:

$$\begin{cases}\mathbf{c}_{j,0}=\frac{1}{N}\sum_{k=1}^{N}\mathbf{x}_k\\ \mathbf{S}_{j,0}=\mathbf{E}^{n\times n},\end{cases}$$

where $\mathbf{E}^{n\times n}$ is identity matrix $n\times n$.

1) At the t-th step, two equations are successively solved:

A) For each $j=1,\dots,K$, the following vector equation is first solved to find $\mathbf{c}_{j,t+1}$:

$$\mathbf{c}_j=\frac{1}{V_j}\sum_{k\in\mathbf{I}_j}v_k\,\mathbf{x}_k,$$

where $z_k=D_k(\mathbf{c}_1,\dots,\mathbf{c}_K;\mathbf{S}_{1,t},\dots,\mathbf{S}_{K,t})$.

B) Then, for each $j=1,\dots,K$, the following vector equation is solved to find $\mathbf{S}_{j,t+1}$:

$$\mathbf{S}_j=\frac{1}{V_j}\sum_{k\in\mathbf{I}_j}v_k\,(\mathbf{x}_k-\mathbf{c}_{j,t+1})'(\mathbf{x}_k-\mathbf{c}_{j,t+1}),$$

where $z_k=D_k(\mathbf{c}_{1,t+1},\dots,\mathbf{c}_{K,t+1};\mathbf{S}_1,\dots,\mathbf{S}_K)$.

2) Step 1 is repeated until $t<T$ (maximum number of iterations) or the sequence $\{Q(\mathbf{c}_{t,1},\dots,\mathbf{c}_{t,K};\mathbf{S}_{t,1},\dots,\mathbf{S}_{t,K})\}$ will not concentrate around its condensation point.

The sets of point indices $\mathbf{I}_1,\dots,\mathbf{I}_K$ corresponding to the partition into clusters are found before solving systems of equations. An additional condition $|\mathbf{S}|=1$ is usually added to prevent singularity of the covariance matrices. Scale factor $\sigma=|\mathbf{S}|$ can then be estimated using the S-estimator [9].

The first equation in the system has the form:

$$\mathbf{c}=F(\mathbf{c}).$$

To solve it, you can use the iterative procedure:

$$\mathbf{c}_{t+1}=(1-h)\mathbf{c}_t+hF(\mathbf{c}_t),$$

where $0\leq h\leq 1$. The second equation has a similar form:

$$\mathbf{S}=G(\mathbf{S}).$$

To solve it, you can use a similar iterative procedure:

$$\mathbf{S}_{t+1}=(1-h)\mathbf{S}_t+hG(\mathbf{S}_t).$$

VI. EXAMPLES

A. IRIS dataset

Consider the relatively simple and classic IRIS dataset (3 classes, 4 attributes, 150 items). Here we use data in projection on 1st and 2nd principial components. As a rule, it is used for classification tasks. Here we will try to identify classes using clustering, using the Mahalanobis distance instead of Euclidean. Figure 1 shows the results of clustering using the robust algorithm proposed here and the classical algorithm. The result of clustering using the robust algorithm (both using $\mathsf{WM}_{\rho_\alpha}$ and $\mathsf{TM}_{\rho_\alpha}$, $\varepsilon = 0.001$, $\alpha = 0.96$) differs from the given classification only at 3 points out of 150. The result of clustering using the classical algorithm differs from the given classification in no less than 6 points out of 150. For comparison, the classic kmeans with Euclidean distance differs from the given classification at 17 points out of 150. This simple example shows that the application of the proposed robust approach to clustering based on a realistic set of features can allow us to construct a partition that differs slightly from a given classification.

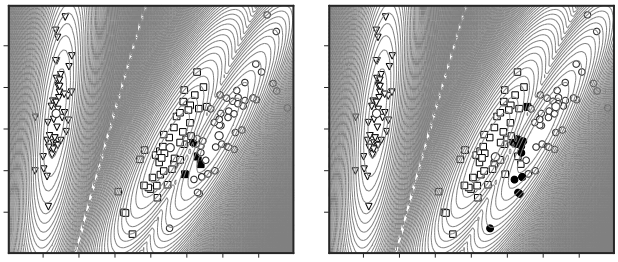

Fig. 1. IRIS: Robust vs. regular algorithm. White markers correspond to samples with the correct classes, black markers correspond to the wrong classes.

B. Wine dataset

Consider another classic WINE dataset (3 classes, 13 attributes, 178 items). As a rule, it is also used for classification tasks. Here we will also try to identify classes using clustering, using the Mahalanobis distance instead of Euclidean. The result of clustering using the robust algorithm (both using $\mathsf{WM}_{\rho_\alpha}$ and $\mathsf{TM}_{\rho_\alpha}$, $\varepsilon = 0.001$, $\alpha = 0.97$) differs from the given classification only at 3–4 points out of 178. The result of clustering using the classical algorithm differs from the given classification in no less than 6–7 points out of 178. This simple example shows that the application of the proposed robust approach to clustering based on a realistic set of features can allow us to construct a partition that differs slightly from a given classification.

C. S1–S4 datasets

Cosider datasets S1–S4 for clustering from . They contain 5000 points, 15 clusters. In Fig. 2–5 presents the results of clustering for sets S1–S4, respectively. On each figure, on the left side there is the result of the robust algorithm, and on the right side there is the classical one. During the training, a robust mean estimate was used with $\mathsf{TM}_{\rho_\alpha}$, $\varepsilon = 0.001$, $\alpha = 0.96 - 0.97$, $h = 0.95$. It is easy to see that the robust algorithm allows one to find more adequate positions of the centers of clusters and the shape of the variance matrices.

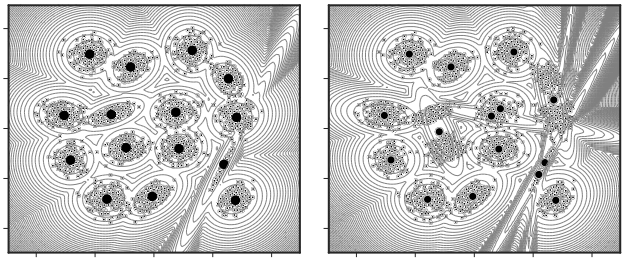

Fig. 2. S1: The results of robust and classical algorithms.

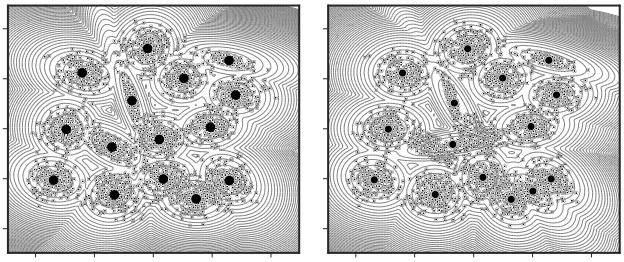

Fig. 3. S2: The results of robust and classical algorithms.

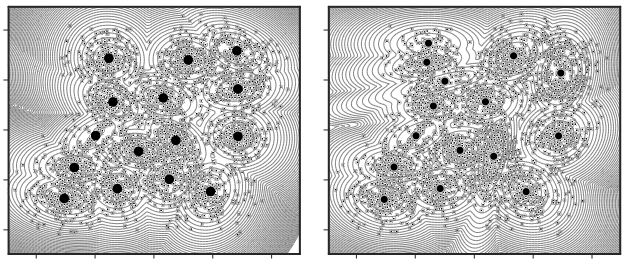

Fig. 4. S3: The results of robust and classical algorithms.

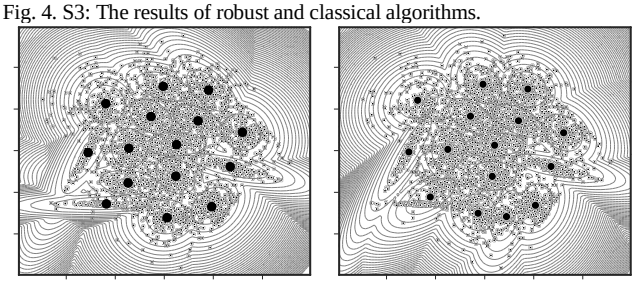

Fig. 5. S4: The results of robust and classical algorithms.

VII. CONCLUSION

In this paper, we considered a new variant of the k-means algorithm, in which the Mahalanobis distance was used instead of the Euclidean distance. The proposed new approach to constructing a robust version of k-means algorithm with the Mahalanobis distance bases on minimizing robust differentiable estimates of the mean. Its fundamental resistance ability to strong distortions in data was shown compared with the classical k-means algorithm. This is due to the fact that the robust average estimates used in the work limit the influence on the search for the position of the centers of clusters of points that are located at relatively large distances from them. The differentiability of the estimate of the average value, insensitive to outliers, allows the use of gradient minimization algorithms to search for cluster centers. Differentiability made it possible to construct an algorithm based on the iterative reweighting method, so that at each step the centers of the clusters are searched within the framework of the classical k-means with sample weights. Taking into account the shape of the covariance matrix significantly enhance the result. It should also be noted that the result of the robust algorithm is not completely stable. However, with a suitable choice of

parameters α and h, it can be achieved that in most starts of the training procedure, an adequate result can be obtained.

ACKNOWLEDGMENT

This work was supported by the RFBR grant No. 18-01-00050-a.

REFERENCES

[1] V.V. Tatarnikov, I.A. Pestunov and V.B. Berikov, "Centroid Averaging Algorithm for Building a Cluster Ensemble," Computer Optics, vol. 41, no. 5, pp. 712-718, 2017. DOI: 10.18287/2412-6179-2017-41-5-712-718.

[2] A.K. Jain, "Data clustering: 50 years beyond K-means," Pattern Recognition Letters, vol. 31, no. 8, pp. 651-666, 2010. DOI: 10.1016/j.patrec.2009.09.011.

[3] S. Belim and P. Kutlunin, "Boundary extraction in images using a clustering algorithm," Computer Optics, vol. 39, no. 1, pp. 119-124, 2015. DOI: 10.18287/0134-2452-2015-39-1-119-124.

[4] R.A. Maronna, "Robust M-Estimators of Multivariate Location and Scatter," Ann. Statist, vol. 4, pp. 51-67, 1976.

[5] P. Rousseeuw and M. Hubert, "High-breakdown estimators of multivariate location and scatter," Robustness and Complex Data Structures, Springer, pp. 49-66, 2013.

[6] R.A. Maronna and V.J. Yohai, "Robust and efficient estimation of multivariate scatter and location," arXiv: 1504.03389, 2015.

[7] Z.M. Shibzukhov, "On the principle of empirical risk minimization based on averaging aggregation functions," Dokl. Math., vol. 96, no. 3, pp. 494-497, 2017. DOI: 10.1134/S106456241705026X.

[8] Z.M. Shibzukhov and M.A. Kazakov, "Clustering based on the principle of finding centers and robust averaging functions of aggregation," Journal of Physics: Conference Series, vol. 1368, 052010, 2019. DOI: 10.1088/1742-6596/1368/5/052010

[9] P.L. Davies, "Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices," Ann. Statist, vol. 15, pp. 1269-1292, 1987.

[10] P. Fränti and S. Sieranoja, "K-means properties on six clustering benchmark datasets," Applied Intelligence, vol. 48, no. 12, pp. 4743-4759, 2018.

[11] Clustering basic benchmark [Online]. URL: http://cs.joensuu.fi/sipu/datasets/.