CLUSTERING AROUND *K* HYPERPLANES WITH THE USE OF CPL CRITERION FUNCTIONS

Leon Bobrowski

[1] Faculty of Computer Science, Bialystok Technical University
[2] Institute of Biocybernetics and Biomedical Engineering, PAS, Warsaw, Poland

The K-means algorithm plays a basic role in non-hierarchical clustering of data sets [1,2]. In this approach K subsets (clusters) C_k are enhanced from a given data set C ($C_1 \cup \ldots \cup C_K = C; C_k \cap C_l = \phi$, $if\ k \neq l$) during a multistage process. The final clusters C_k result from a sequence of current clusters $C_k[l]$ modifications ($l = 1, 2, \ldots$). The current clusters $C_k[l]$ are defined during the l-th stage of the clustering process. The number K of clusters $C_k[l]$ is usually fixed at the beginning of the clustering process.

Each stage of the clustering process consists of two steps. During the first step the centers ("means") $\mathbf{m}_k[l]$ of current clusters $C_k[l]$ are computed. During the next stage, the membership of all elements (feature vectors) $\mathbf{x}_j$ of each current cluster $C_k[l]$ is verified and possibly changed in accordance with the principle of the nearest center. The principle of the nearest center means that each element $\mathbf{x}_j$ of the data set C ($\mathbf{x}_j \in C$) is classified to this cluster $C_k[l]$ for which the distance $\rho(\mathbf{x}_j, \mathbf{m}_k[l])$ between the vector $\mathbf{x}_j$ and the center $\mathbf{m}_k[l]$ is minimal. Replacement of the set $C_k[l]$ by $C_k[l+1]$ usually results in some modification of the center $\mathbf{m}_k[l]$. The outcome of the K-means procedure depends among others on the type of the distance function $\rho(\mathbf{x}_j, \mathbf{m}_k[l])$ used in finding of the nearest (the most similar) center $\mathbf{m}_k[l]$. Both the Euclidean distance $\rho_E(\mathbf{x}_j, \mathbf{m}_k[l])$ as well as the non-Euclidean measures of similarity can be used in the K-means algorithm [6].

The presented K-plans algorithm can be treated as some kind of modification of the K-means algorithm. The centers $\mathbf{m}_k[l]$ of current clusters $C_k[l]$ are replaced by the central hyperplanes $H(\mathbf{w}_k[l], \theta_k[l])$

$$H(\mathbf{w}_k[l], \theta_k[l]) = \{\mathbf{x} : \mathbf{w}_k[l]^T \mathbf{x} = \theta_k[l]\} \ . \tag{1}$$

where $\mathbf{x}$ is the feature vector ($\mathbf{x} \in R^n$), $\mathbf{w}_k = [w_{k1}, \ldots, w_{kn}]^T \in R^n$ is the weight vector, $\theta_k \in R^1$ is the threshold, and $(\mathbf{w}_k[l])^T \mathbf{x}$ is the inner product.

The distance $\rho(\mathbf{x}_j; \mathbf{w}_k[l], \theta_k[l])$ of the feature vector $\mathbf{x} - j$ from the hyperplanes $H(\mathbf{w}_k[l], \theta_k[l])$ can be computed in the following manner:

$$\rho(\mathbf{x}_j; \mathbf{w}_k[l], \theta_k[l]) = |\mathbf{w}_k[l]^T \mathbf{x}_j / \|\mathbf{w}_k[l]\| - \theta_k[l] / \|\mathbf{w}_k[l]\|| \tag{2}$$

The central hyperplane $H(\mathbf{w}_k[l], \theta_k[l])$ (1) should represent the current cluster $C_k[l]$. Such hyperplanes can be defined through the minimization of the convex and piecewise linear (CPL) criterion functions $\Phi_k(\mathbf{w})$ [6]:

$$\Phi_k(\mathbf{w}) = \sum_{j \in J_k} \alpha_j \varphi_j(\mathbf{w}), \quad where \quad k = 1, \ldots, K \tag{3}$$

where J_k is the set of indices j of such feature vectors $\mathbf{x}_j$ which belong to the current cluster $C_k[l]$ and $\varphi_j(\mathbf{w})$ is the penalty function related to the feature vector $\mathbf{x}_j$:

$$(\forall \mathbf{x}_j \in C) \quad \varphi_j(\mathbf{w}) = \begin{matrix} \delta - \mathbf{w}^T \mathbf{x}_j & if & \mathbf{w}^T \mathbf{x}_j \leq \delta \\ \mathbf{w}^T \mathbf{x}_j - \delta & if & \mathbf{w}^T \mathbf{x}_j > \delta \end{matrix} \tag{4}$$

[1] The work was partially financed by the KBN grant 3T11F01130, and by the grant S/WI/2/08 from the Białystok Technical University.

where δ is some parameter (margin) ($\delta \in R$). The positive parameters α_j in the functions $\Phi_k(\mathbf{w})$ (3) and can be treated as the prices of particular feature vectors $\mathbf{x}_j$.

The basis exchange algorithms which are similar to linear programming allow to find the minimum of the criterion functions $\Phi_k(\mathbf{w})$ (3) efficiently, even in the case of large, multidimensional sets $C_k[l]$ [5].

$$(\exists \mathbf{w}_k^*) \quad (\forall \mathbf{w}) \quad \Phi_k(\mathbf{w}) \leq \Phi_k(\mathbf{w}_k^*) = \Phi_k^* \tag{5}$$

It can be proved that the minimal value Φ_k^* of the criterion functions $\Phi_k(\mathbf{w})$ (3) is equal to zero ($\Phi_k^* = 0$) if and only if all the feature vectors $\mathbf{x}_j$ from the sets $C_k[l]$ can be situated on some hyperplane $H(\mathbf{w}_k[l], \theta_k[l])$ (1) with $\theta = 0$ [].

The optimal vectors $\mathbf{w}_k^*$ (5) constitute the minimal values Φ_k^* of the criterion functions $\Phi_k(\mathbf{w})$ (3) which are defined on the current clusters $C_k[l]$. Each vector $\mathbf{w}_k^*$ (5) also allows to determine the central hyperplane $H(\mathbf{w}_k[l], \delta)$ (1) of the set $C_k 1[l]$. In the next stage clusters $C_k[l+1]$ can be defined in accordance with the principle of the nearest central hyperplane $H(\mathbf{w}_k^*, \delta)$. The set $C_k[l+1]$ contains such feature vectors $\mathbf{x}_j$ for which the distance $\rho(\mathbf{x}_j; \mathbf{w}_k^*, \delta)$ (2) between the vector $\mathbf{x}_j$ and the central hyperplane $H(\mathbf{w}_k[l], \delta)$ is the minimal one. In this way, the iterative procedure similar to the K-means algorithm can be implemented.

References

[1] Duda O. R, Hart P. E., Stork D. G.: Pattern Classification, J. Wiley, New York, 2001.
[2] Fukunaga K.: Introduction to Statistical Pattern Recognition, Academic Press 1990.
[3] Bobrowski L., Bezdek J. C., "C-means clustering with the L_1 and L_∞ norms", IEEE Transactions on Systems Man and Cybernetics, Vol. 21, No. 3, pp. 545-554, May/June 1991.
[4] Bobrowski L.: “CPL clustering with feature costs”, ICDM2008, Leipzig, Germany, 2008
[5] Bobrowski L.: "Design of piecewise linear classifiers from formal neurons by some basis exchange technique’" Pattern Recognition, 24(9), pp. 863-870, 1991.
[6] Bobrowski L.: Eksploracja danych oparta na wypukłych i odcinkowo-liniowych funkcjach kryterialnych (Data mining based on convex and piecewise linear (CPL) criterion functions) (in Polish), Białystok Technical University, 2005.