Fuzzy C-ordered means clustering algorithm

Fuzzy clustering helps to find natural vague boundaries in data. However, this method's sensitivity to noise and outliers in the data is one of its biggest drawbacks. Numerous robust loss function proposals, including the Huber's (HUB), logarithmic (LOG), and sigmoidal (SIG) loss functions, are employed in the studies to get around this problem. Fuzzy C-Ordered-Means (FCOM) clustering is a robust fuzzy clustering technique that was introduced by Leski (2016). To achieve its robustness, this method makes use of both Huber's M-estimators and Yager's OWA operators. The objective function of the FCOM algorithm is defined as follow:

|  |  |
| --- | --- |
|  | (1) |

Subject to

|  |  |
| --- | --- |
|  | (2) |

In eq. (1), denotes the typically of the *n*th point for the *k*th cluster; a smaller corresponds to more atypical data. These parameters are derived based on the ordering the distances of data from prototypes. The parameter can be interpreted as an overall (general) typicality of the *n*th datum, which depends on typicality of the *n*th datum with respect to all the clusters, and the formula of can be described as:

|  |  |
| --- | --- |
|  | (3) |

The loss function is used to measures the distance. Therefore, is represented as follows:

|  |  |
| --- | --- |
|  | (4) |

By applying the Lagrange multiplier approach, the following updating equations are derived:

|  |  |
| --- | --- |
|  | (5) |

|  |  |
| --- | --- |
|  | (6) |

|  |  |
| --- | --- |
|  | (7) |

where parameters depend on the loss function used and residuals . The center of cluster, , depends on parameters, but the values of depend on the obtained residuals . In turn, the residuals depend on . Consequently, in order to minimize the objective function, it is necessary to iteratively re-weight the ordering of the residuals. Therefore, the objective function should be minimized by iteratively re-weighting the ordering of the residuals.

The ordering of the residuals indicates the distance from the data point to the center of cluster requests to be discussed. Each feature of data points obtains the rank-ordered by the residuals to calculate the typically. Let denotes typically of the *m*th feature of *n*th point with respect to the *k*th cluster. The represents the rank-ordered from to . The smaller the means the smaller the residual. The parameter can be calculated from the following two equations:

|  |  |
| --- | --- |
|  | (8) |

or

|  |  |
| --- | --- |
|  | (9) |

Eqs. (8) and (9) are called Piecewise-Linearly-weighted OWA (PLOWA) and Sigmoidally-weighted OWA (SOWA), respectively. The parameter which denotes typically of the *n*th point with respect to the *k*th cluster can be defined as:

|  |  |
| --- | --- |
|  | (10) |

Some loss functions are expressed as follow:

|  |  |
| --- | --- |
| Loss functions:   * LINear (LIN) error function * HUBer with parameter * SIGmoidal (SIG) with parameters * SIGmoidal-Linear (SIGL) with parameters * LOGarithmic (LOG) * LOG-Linear (LOGL) |  |