Fuzzy Weighted C-ordered means clustering algorithm

The fuzzy C-ordered means algorithm (FCOM) [Leski, 2016] calculates typicality of each data item. The data items are ordered and their typicality is updated in each iteration of the clustering procedure. However, this algorithm does not assign weights to attributes. The Fuzzy Weighted C-Ordered-Means (FWCOM) clustering proposed by Siminski in 2017 handles this drawback. This algorithm finds clusters in fuzzy subspaces of the original task space and assigns weights to features in each cluster. The weights are numbers from unit interval [0, 1]. To make this algorithm more robust to outliers and noise it incorporates the ordering technique. This algorithm minimizes a criterion function defined as:

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

Subject to

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

In eq. (1), denotes the typically of the *n*th point for the *k*th cluster; is a by matrix, where refers to the weight of *m*th feature in *k*th cluster, is a weighting exponent for weights. Additionally, is a global typically of the *n*th data point, which depends on typicality of the *n*th data point with respect to all the clusters, and can be described as:

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

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

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

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

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

In eq. (6), is a residual of the *n*th data from the center of the *k*th cluster, and is a loss function defined in various ways:

|  |  |
| --- | --- |
| * Essential loss function * Absolute (Linear) loss function * HUBer with parameter * SIGmoidal (SIG) with parameters * SIGmoidal-Linear (SIGL) with parameters * LOGarithmic (LOG) * LOG-Linear (LOGL) | (7)  (8)  (9)  (10)  (11)  (12)  (13) |

For each data item in each cluster the typicality is calculated for each feature separately. In order to do this data items are ordered by their distances from the cluster center. The closest data item with respect to the *m*th feature is labelled with an ordinal number1, the most distant with (where is the number of data items). The value denotes typicality of the *n*th data item with respect to the *k*th cluster and *m*th feature. This parameter can be calculated from the following two equations:

|  |  |
| --- | --- |
|  | (14) |

or

|  |  |
| --- | --- |
|  | (15) |

Eqs. (14) and (15) are called Piecewise-Linearly-weighted OWA (PLOWA) and Sigmoidally-weighted OWA (SOWA), respectively. In both functions stands for the index of the *n*th data item after reordering by the distance from the *k*th cluster with respect to the *m*th feature. The parameter which denotes typically of the *n*th point with respect to the *k*th cluster can be defined as:

|  |  |
| --- | --- |
|  | (16) |