Lesson 4. Clustering algorithm in real life problems

K-means clustering

The algorithm randomly assigns each observation into one of k categories, then calculates the mean of each category. Next, it reassigns each observation to the category with the closest mean before recalculating the means. This step repeats over and over until no more reassignments are necessary.

ISODATA clustering

ISODATA unsupervised classification calculates class means evenly distributed in the data space then iteratively clusters the remaining pixels using minimum distance techniques. Each iteration recalculates means and reclassifies pixels with respect to the new means.

Hierarchical clustering

The algorithm is based on distance recalculation between clusters. Clusters with the smallest distance between each other are merged until the number of clusters reaches the desired number. This method uses dendrograms.

DBSCAN

The number of clusters is not required. Can find clusters of stable forms. It may even find a cluster that is completely surrounded by, but not connected to, another cluster. Creates noise and is resistant to emissions. Only two parameters required

How to use?

Open the group of clustering tools  *-> select the desired method -> set the parameters for the distance or number of clusters -> select the data on the basis of which clustering will be carried out (spatial, attributive or both) ->Evaluate visual result, change inputparameters if result is unsatisfactory*