chapter-clustering-validation

Ar

2022-07-19

Need for validation of clusters

Clustering algorithms are designed such that they come out with a given number of clusters even if the underlying data is devoid of any such clusters. We will see a criterion to assess the credibility of the clusters produced by any clustering algorithm.

Within-groups sum of squared distances (WSS):

$$WSS_k = \sum_{l=1}^{k} \sum_{x_i \in C_l} d^2(x_i, \overline{x_l})$$

where, k is the number of clusters and within each cluster l, x_l is the centre of mass.