

Davies Bouldin Index

Clustering is an important class of unsupervised learning. Clustering algorithms cluster(or group) data points based on the similarity between them. This grouping is different from classification as unlike classification, groups are not predefined. There are different types of clustering algorithms - Exclusive Clustering, Overlapping Clustering, Hierarchical Clustering, Probabilistic Clustering, etc.

Unlike classification, in which measuring quality of classification is very straightforward, quantifying clustering quality requires setting up similarity measures whose computation can sometimes become too complex! *Davies Bouldin Index* is one such measure of computing the quality of clustering that has been performed.

The Davies Bouldin Index has to be calculated for any value of n_clusters (nc) as follows:

$$DB = \frac{1}{n_c} \sum_{i=1}^{n_c} R_i, \text{ where}$$

$$R_i = \max_{j=1 \dots n_c, i \neq j} (R_{ij}), \quad i = 1 \dots n_c$$

$$R_{ij} = \frac{s_i + s_j}{d_{ij}}$$

$$d_{ij} = d(v_i, v_j), \quad s_i = \frac{1}{\|c_i\|} \sum_{x \in c_i} d(x, v_i)$$

Where,

- $d(x,y)$ is the Euclidean distance between x and y .
- c_i is the cluster i .
- v_i is the centroid of cluster c_i
- $\|c_i\|$ refers to the norm of c_i

(Davies Bouldin Index: https://en.wikipedia.org/wiki/Davies%E2%80%93Bouldin_index)

(K-means clustering using scikit-learn: http://scikit-learn.org/stable/tutorial/statistical_inference/unsupervised_learning.html)

Input format

Line 1: M values

Line 2: N features

Line 3: $n_clusters$

Next M lines contain N space separated floating point values.

Line 4 to $M+3$: $v_1 v_2 \dots v_N$

Output format

Output a single integer, the *Davies Bouldin Index* for the given Input when k-means clustering algorithm is applied to it with given number of cluster centers.

Result:

RESULT:  Accepted

Score	Time (sec)	Memory (KiB)	Language
10.0	0.89335	50440	Python 3