# Silhouette coefficient

## Definition and purpose

Silhouette refers to a method of interpretation and validation of consistency within clusters of data. The technique provides a succinct graphical representation of how well each object has been classified.

Assume the data has been clustered into groups. For data point let

Which referred the **intracluster** **distance**. And then we define the mean dissimilarity (**intercluster distance**)

And

For to be close to 1 we require . **As is a measure of how dissimilar is to its own cluster**, a small value means it is well matched. Furthermore, a large implies that is badly matched to its neighbouring cluster. Thus **an close to 1 means that the data is appropriately clustered**. **If is close to -1, then by the same logic we see that would be more appropriate if it was clustered in its neighboring cluster**. **An near zero means that the datum is on the border of two natural clusters**.