

Glossary of machine learning terms

Click on each section title to see the Wikipedia entry for that term.

clustering

Blah

distance metric

Blah

k-means

Blah

silhouette score

After *clusters* have been defined, the *silhouette score* for each element quantifies how similar that element is to the other elements in the same cluster, relative to the elements belonging to other clusters. This is calculated on a scale from -1 to +1, and requires us to choose a suitable *distance metric*.

By summing the silhouette scores for all elements, we can produce an overall score for the clustering, which can be compared to the scores obtained from other potential clustering solutions (for example, those obtained by using different values of k in *k-means*). By this measure, the best clustering will have the largest silhouette score.