

Elbow method (clustering)

From Wikipedia, the free encyclopedia

The **Elbow method** is a method of interpretation and validation of consistency within cluster analysis designed to help finding the appropriate number of clusters in a dataset.

This method looks at the percentage of variance explained as a function of the number of clusters: One should choose a number of clusters so that adding another cluster doesn't give much better modeling of the data. More precisely, if one plots the percentage of variance explained by the clusters against the number of clusters, the first clusters will add much information (explain a lot of variance), but at some point the marginal gain will drop, giving an angle in the graph. The number of clusters is chosen at this point, hence the "elbow criterion". This "elbow" cannot always be unambiguously identified.^[1] Percentage of variance explained is the ratio of the between-group variance to the total variance, also known as an F-test. A

slight variation of this method plots the curvature of the within group variance.^[2]

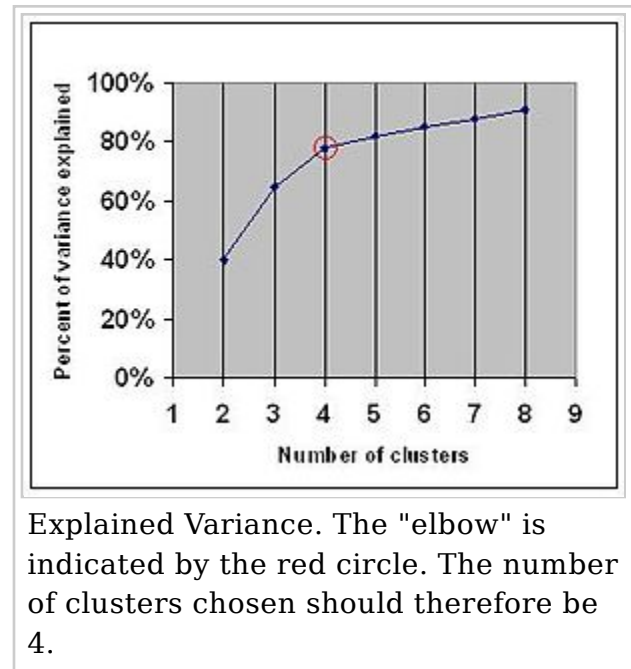
The method can be traced to speculation by Robert L. Thorndike in 1953.^[3]

See also

- Determining the number of clusters in a data set
- Silhouette (clustering)

References

1. See, e.g., David J. Ketchen, Jr; Christopher L. Shook (1996). "The application of cluster analysis in Strategic Management Research: An analysis and critique". *Strategic Management Journal*. **17** (6): 441–458. doi:10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G.
2. See, e.g., Figure 6 in
 - Cyril Goutte, Peter Toft, Egill Rostrup, Finn Årup Nielsen, Lars Kai Hansen (March 1999). "On Clustering fMRI Time Series". *NeuroImage*. **9** (3): 298–310.



doi:10.1006/nimg.1998.0391. PMID 10075900.

3. Robert L. Thorndike (December 1953). "Who Belongs in the Family?". *Psychometrika*. **18** (4): 267–276. doi:10.1007/BF02289263.

Retrieved from "[https://en.wikipedia.org/w/index.php?title=Elbow_method_\(clustering\)&oldid=722750167](https://en.wikipedia.org/w/index.php?title=Elbow_method_(clustering)&oldid=722750167)"

Categories: Clustering criteria | Computer science stubs

- This page was last modified on 29 May 2016, at 23:23.
- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.