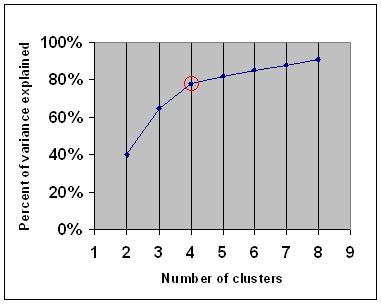
**ELBOW METHOD**

The [**Elbow Method**](https://en.wikipedia.org/wiki/Elbow_method_(clustering)) 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 modelling of the data. 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. Percentage of variance explained is the ratio of the between-group variance to the total variance, also known as an [F-test](https://en.wikipedia.org/wiki/F-test). A slight variation of this method plots the curvature of the within group variance.



*Fig.6 Elbow method graph*

“How might the algorithm be modified to diminish such sensitivity?” Instead of taking the mean value of the objects in a cluster as a reference point, we can pick actual objects to represent the clusters, using one representative object per cluster. Each remaining object is clustered with the representative object to which it is the most similar. The partitioning method is then performed based on the principle of minimizing the sum of the dissimilarities between each object and its corresponding reference point. That is, an absolute-error criterion is used, defined as

E = k ∑ j=1 ∑ p∈Cj |p−oj |

where E is the sum of the absolute error for all objects in the data set;

p is the point in space representing a given object in clusterCj ; a

nd oj is the representative object ofCj .

In general, the algorithm iterates until, eventually, each representative object is actually the medoid, or most centrally located object, of its cluster. This is the basis of the k-medoids method for grouping n objects into k clusters

Algorithm: k-medoids. PAM, a k-medoids algorithm for partitioning based on medoid or central objects.

Input: k: the number of clusters, D: a data set containing n objects.

Output: A set of k clusters.

Method: (1) arbitrarily choose k objects in D as the initial representative objects or seeds;

(2) repeat

(3) assign each remaining object to the cluster with the nearest representative object;

(4) randomly select a nonrepresentative object, orandom;

(5) compute the total cost, S, of swapping representative object, oj , with orandom;

(6) if S < 0 then swap oj with orandom to form the new set of k representative objects;

(7) until no change;

SSE value for given k: = 98.452678

k = 2

SSE value for given k: = 69.379355

k = 3

SSE value for given k: = 64.907219

k = 4

SSE value for given k: = 36.124876

k = 5

SSE value for given k: = 27.128494

k = 6

SSE value for given k: = 23.230394

k = 7

SSE value for givenk: = 20.627080

k = 8

SSE value for given k: = 13.245165

k = 9

SSE value for given k: = 7.414214

k = 10

SSE value for given k: = 2.414214

k = 11

SSE value for given k: = 1.000000

|  |  |
| --- | --- |
| 1 | 121 |
| 2 | 79 |
| 3 | 51 |
| 4 | 39 |
| 5 | 31 |
| 6 | 23 |
| 7 | 16 |
| 8 | 12 |
| 9 | 6 |
| 10 | 3 |
| 11 | 1 |

The process of grouping a set of physical or abstract objects into classes of similar objects is called clustering. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters. A cluster of data objects can be treated collectively as one group and so may be considered as a form of data compression. Although classification is an effective means for distinguishing groups or classes of objects,it requires the often costly collection and labeling of a large set of training tuples or patterns, which the classifier uses to model each group. It is often more desirable to proceed in the reverse direction: First partition the set of data into groups based on data similarity (e.g., using clustering), and then assign labels to the relatively small number of groups. Additional advantages of such a clustering-based process are that it is adaptable to changes and helps single out useful features that distinguish different groups.