# Notes 06-6: Terminology

For numeric attributes, clusters can be described by several characteristic values. Assume a cluster *Kb* consisting of *n* *m*-dimensional points $\left\{\left(p\_{11},p\_{12},…,p\_{1m}\right),\left(p\_{21},p\_{22},…,p\_{2m}\right),…\left(p\_{n1},p\_{n2},…,p\_{nm}\right)\right\}$.

The *centroid*, *Ca*, of a cluster *Ka* is the “middle” point of the cluster (it need not be an actual point in the cluster) and is described by $C\_{a}=\left(p\_{1},p\_{2},…,p\_{m}\right)$ , where *pu*, the *u*-th attribute of the centroid, is given by

$$p\_{u}=\frac{\sum\_{i=1}^{n}p\_{iu}}{n}$$

The *radius*, *Ra*, of a cluster *Ka* is the square root of the average mean squared distance from all points in the cluster to the centroid, and is given by

$$R\_{a}=\sqrt{\frac{\sum\_{i=1}^{n}\sum\_{u=1}^{m}\left|p\_{iu}-p\_{u}\right|^{2}}{n}}$$

The diameter, *Diametera*, of cluster *Ka* is the square root of the average mean squared distance between all pairs of points in the cluster, and is given by

$$Diameter\_{a}=\sqrt{\frac{\sum\_{i=1}^{n}\sum\_{j=1}^{n}\sum\_{u=1}^{m}\left|p\_{iu}-p\_{ju}\right|^{2}}{n(n-1)}}$$

Many clustering algorithms require that the *distance between clusters* be determined (as opposed to the *distance between objects*) to identify when two clusters are of sufficient similarity to be linked together (i.e., amalgamated).

The *single linkage* (or *nearest neighbor*) method links clusters when the distance between the two closest objects in the different clusters is below some threshold.

The *complete linkage* (or *furthest neighbor*) method links clusters when the distance between the two furthest objects in the different clusters is below some threshold.

The *pair-group average* method links clusters when the average distance between all pairs of objects in the different clusters is below some threshold.

The *pair-group centroid* method links clusters when the distance between centroids is below some threshold.

The *pair-group medoid* method links clusters when the distance between medoids is below some threshold.