

Clustering

Slides from Eamonn Keogh at UC Riverside

What is Clustering?

Organizing data into classes such that there is:

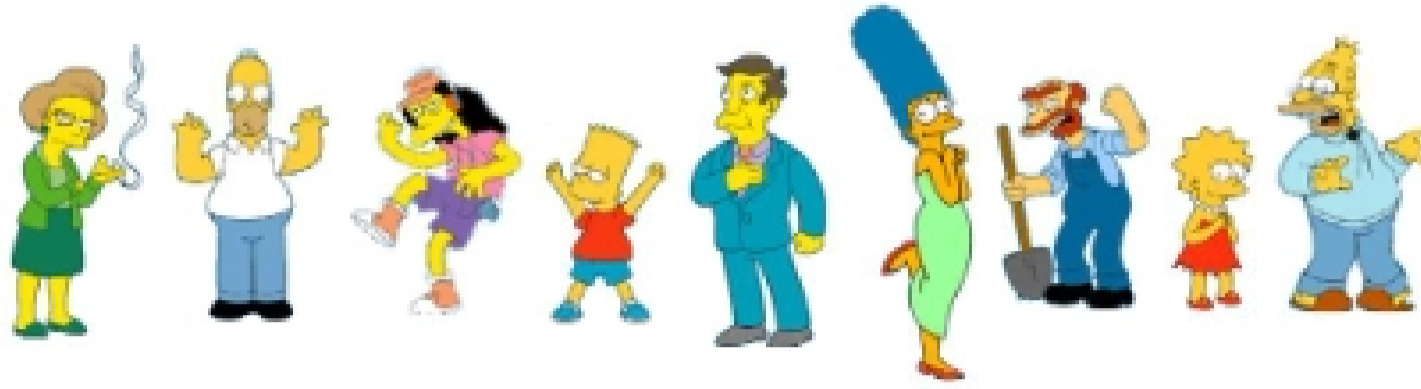
- high intra-class similarity
- low inter-class similarity

More informally, finding natural groupings among objects.

Also called unsupervised learning, sometimes called classification by statisticians, sorting by psychologists, and segmentation by people in marketing

“Unsupervised” as compared to “supervised” learning, where you are given labeled examples (with which you can construct training/test sets)

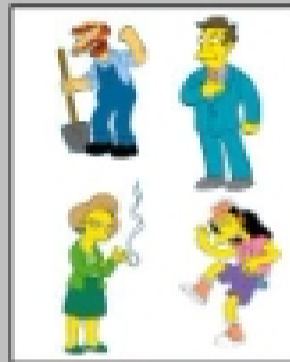
What is a natural grouping?



Clustering is subjective



Simpson's Family



School Employees



Females



Males

How could you cluster these examples? By what traits?