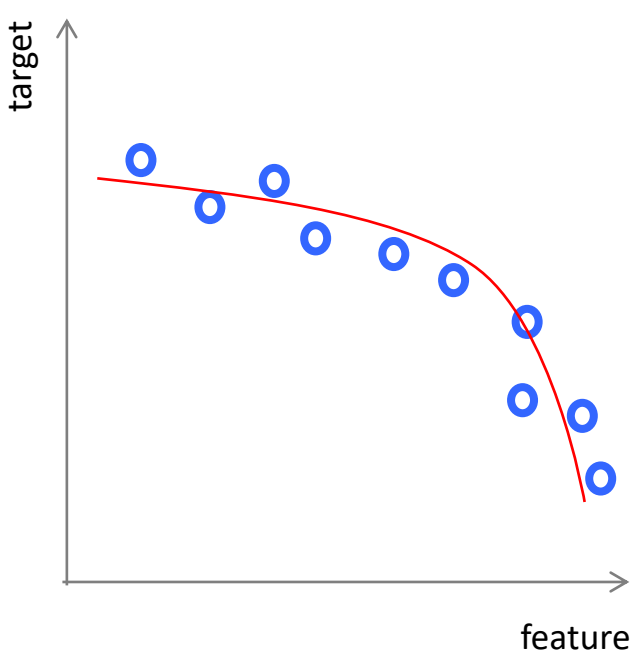
13. Unsupervised Learning: Clustering

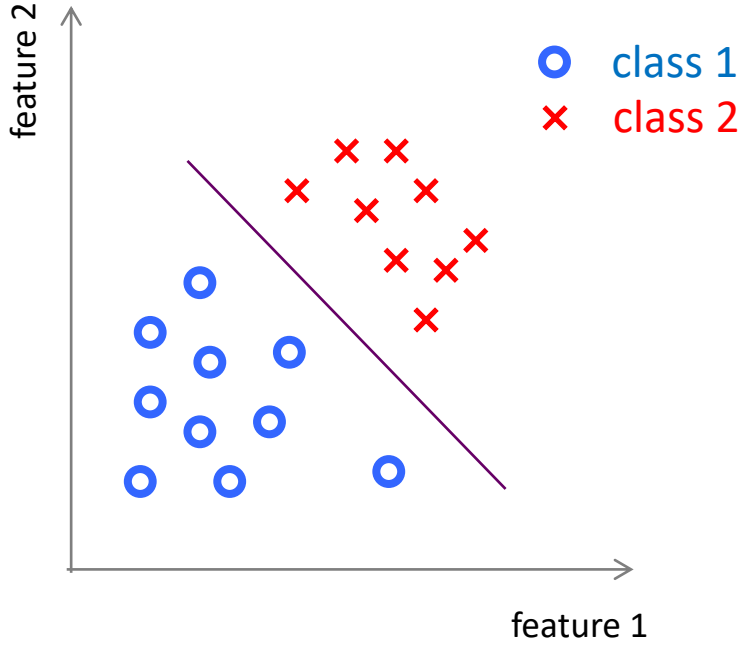
Unsupervised vs Supervised Learning

Supervised Learning: Regression



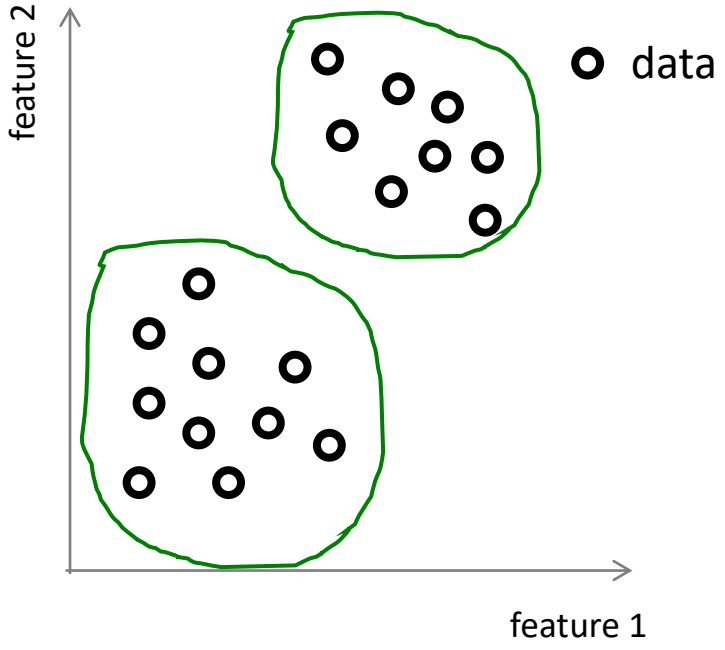
The data is labelled

Supervised Learning: Classification



The data is labelled

Unsupervised Learning: Clustering

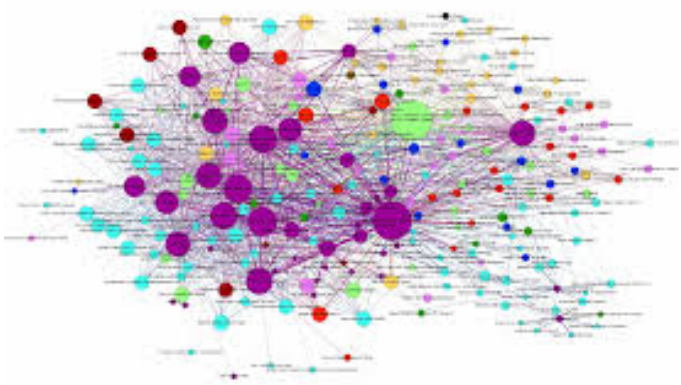


The data is NOT labelled

Unsupervised Learning: Examples

Market segmentation

Social network analysis



Semantic image segmentation

Clustering: What is Clustering?

What is clustering?

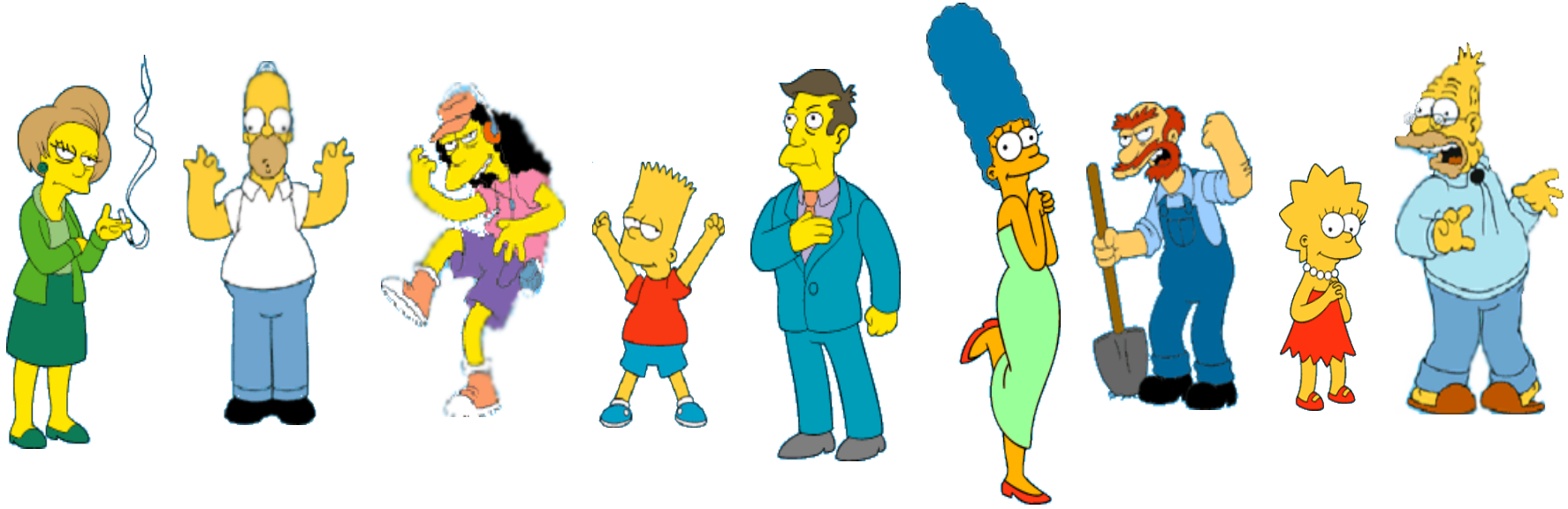
- Clustering involves the organization of the data objects onto ***clusters*** such that there is:
  - high **intra**-cluster similarity (i.e., objects within the same cluster are “close” to one another)
  - low **inter**-cluster similarity (i.e., objects in different clusters are “far” from one another)

- Clustering therefore involves finding the number of clusters and the cluster labels directly from the data (in contrast to classification)

- More informally, clustering involves finding natural groupings among data samples.

Clustering: Toy Example

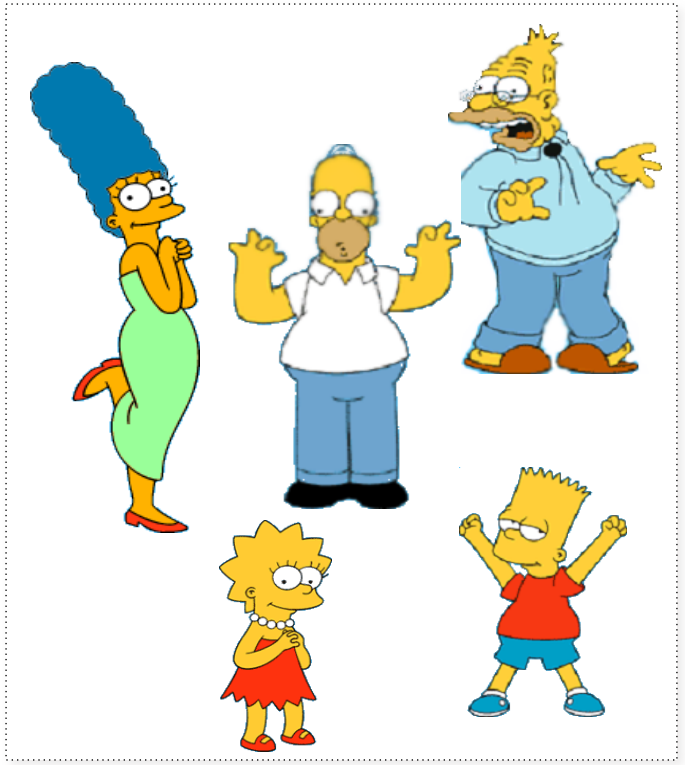
Toy example: Can we group these characters onto meaningful groups?



Clustering: Toy Example

Toy example: Can we group these characters onto meaningful groups?

Cluster 1: Simpson's Family

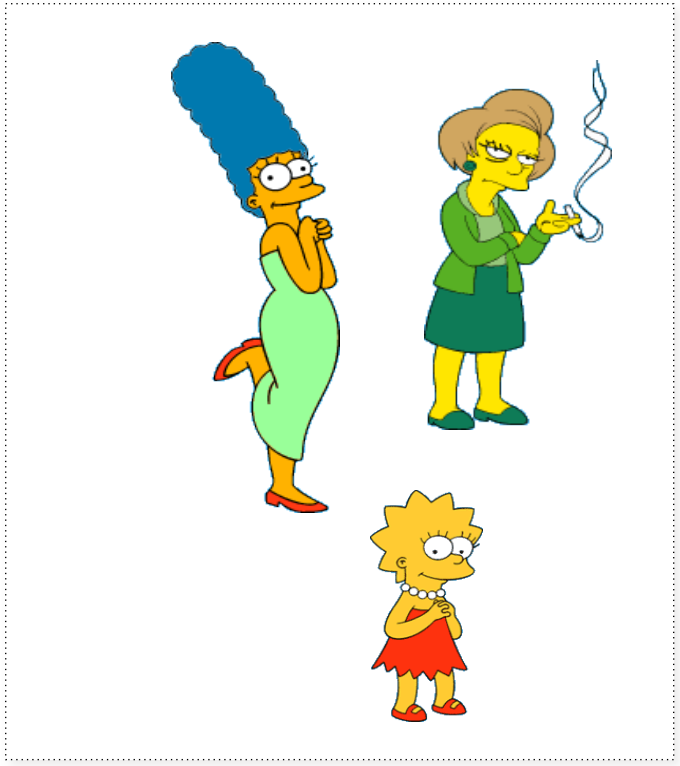


Cluster 2: School Employees

Clustering: Toy Example

Toy example: Can we group these characters onto meaningful groups?

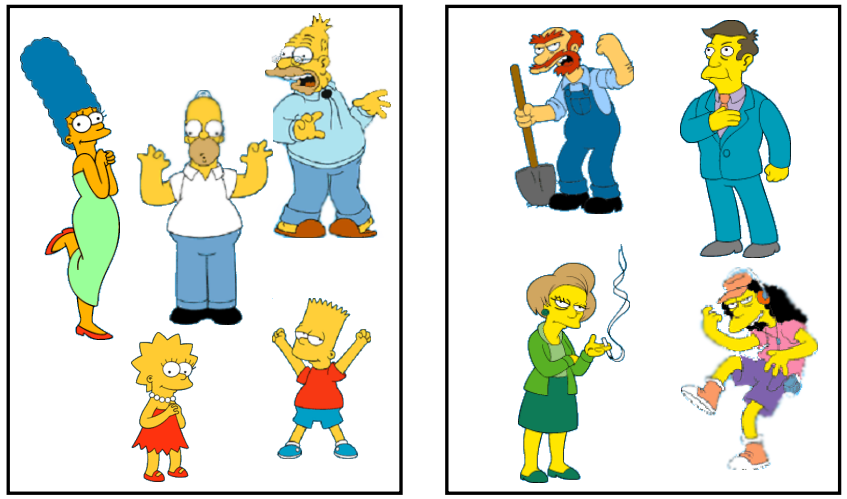
Cluster 1: Females



Cluster 2: Males

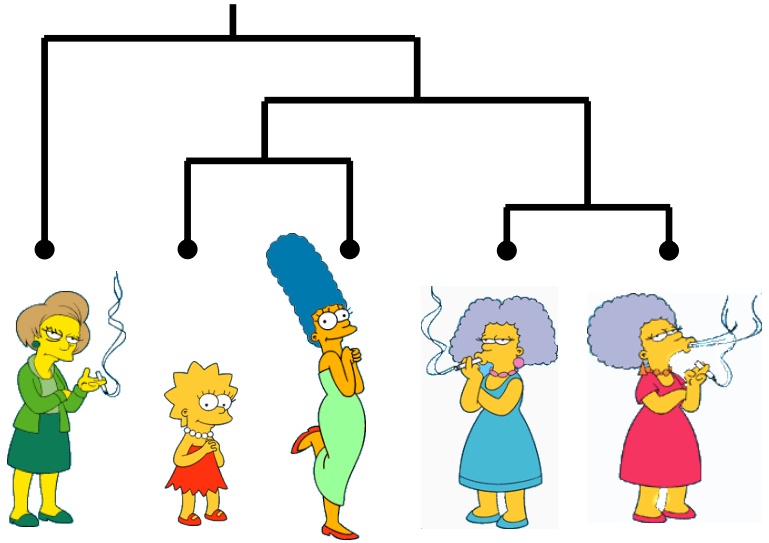
Clustering: Types

Partition Clustering Approaches



Examples: k-means clustering algorithm

Hierarchical Clustering Approaches



Examples: hierarchical clustering algorithms (agglomerative or divisive)