Cluster Analysis

After this video you will be able to..

- Articulate the goal of cluster analysis
- Discuss whether cluster analysis is supervised or unsupervised.
- List some ways that cluster results can be applied

Cluster Analysis Overview

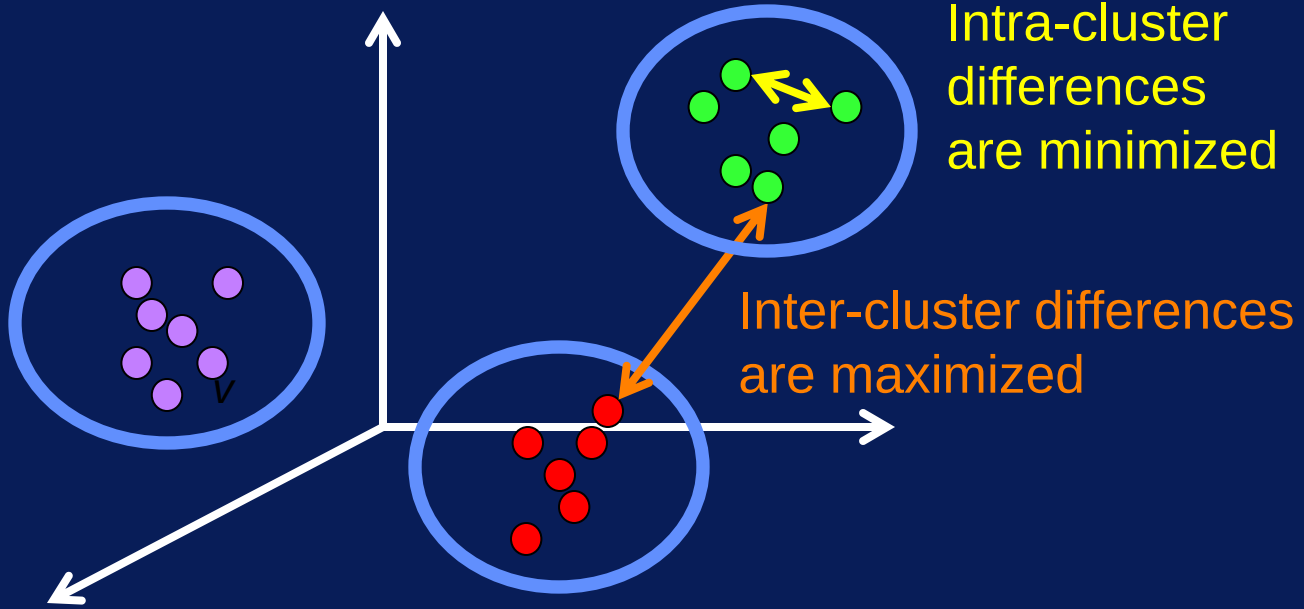
Goal: Organize similar items into groups.

Cluster Analysis Examples

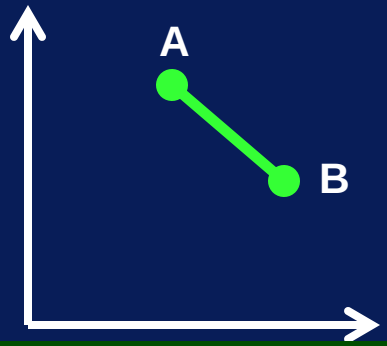
- Segment customer base into groups
- Characterize different weather patterns for a region
- Group news articles into topics
- Discover crime hot spots

Cluster Analysis

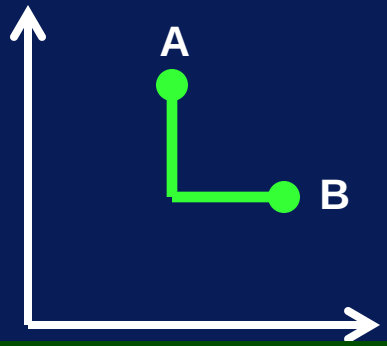
- **Divides data into clusters**
- **Similar items are placed in same cluster**



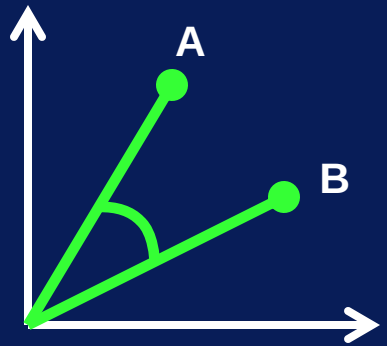
Similarity Measures



Euclidean Distance



Manhattan Distance



Cosine Similarity

Normalizing Input Variables

Weight

Scaled Values

Height

Cluster Analysis Notes

Unsupervised

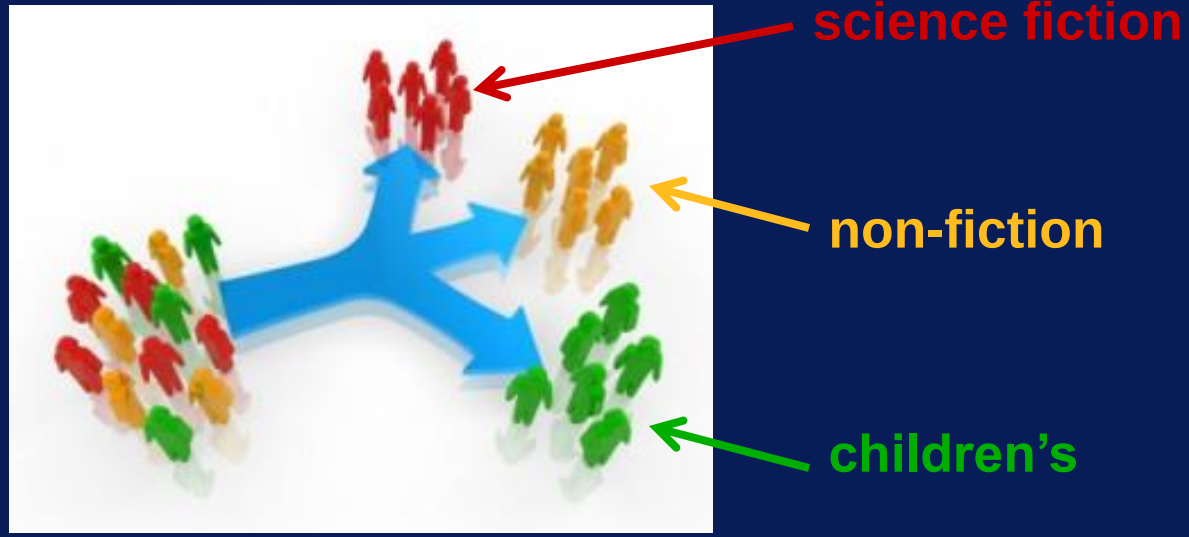
There is no ‘correct’ clustering

Clusters don’t come with labels

Interpretation and analysis required to make sense of clustering results!

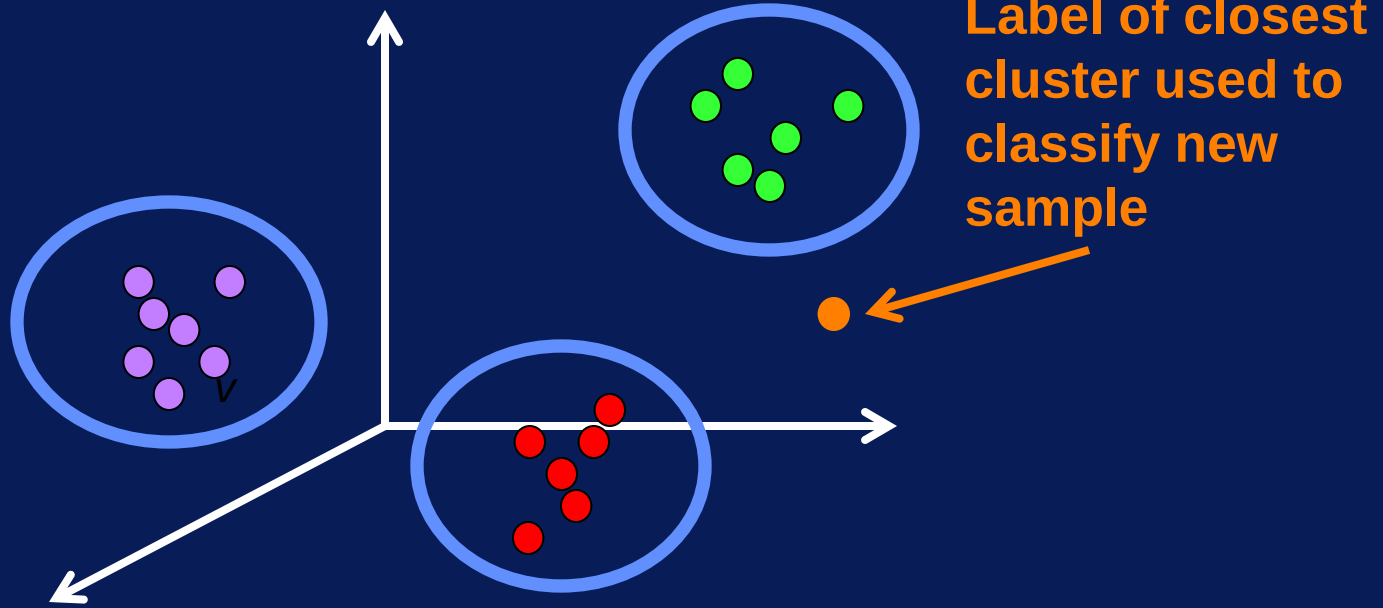
Uses of Cluster Results

- **Data segmentation**
  - Analysis of each segment can provide insights



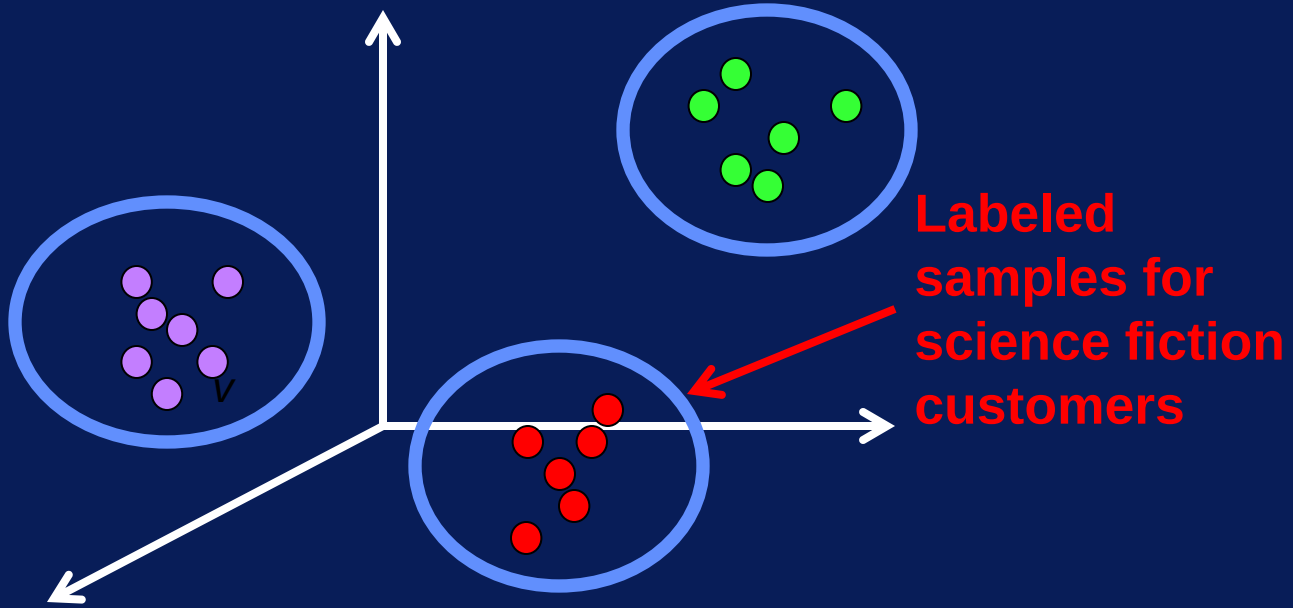
Uses of Cluster Results

- **Categories for classifying new data**
  - New sample assigned to closest cluster



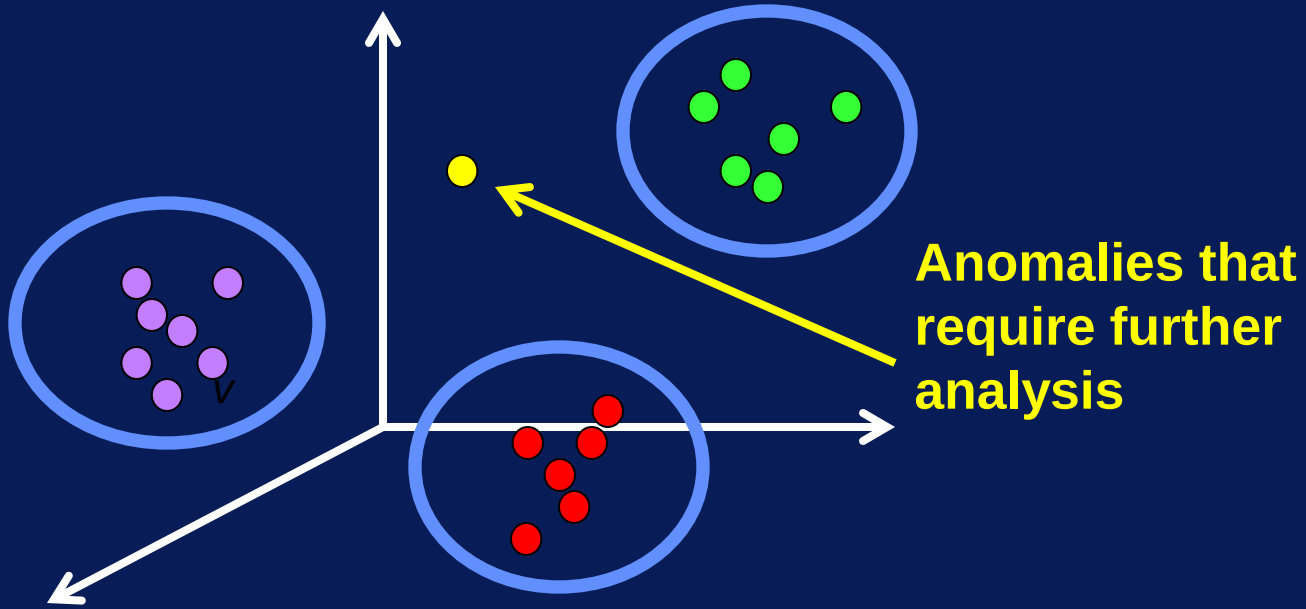
Uses of Cluster Results

- **Labeled data for classification**
  - Cluster samples used as labeled data



Uses of Cluster Results

- **Basis for anomaly detection**
  - Cluster outliers are anomalies



Cluster Analysis Summary

- Organize similar items into groups
- Analyzing clusters often leads to useful insights about data
- Clusters require analysis and interpretation