
DLI Accelerated Data Science Teaching Kit

# Lecture 15.1 - Introduction to Unsupervised Learning

# Supervised Learning vs Unsupervised Learning

**Supervised Learning**

**Data:** (x, y)

**Goal:** Learn function to map, discover patterns in the data that relate data attributes with a target (class) attribute.

**Examples:** Classification, Regression, Object detection, Object tracking, etc.

**Unsupervised Learning**

**Data:** x

**Goal:** Learn the hidden or intrinsic structure of the data

**Examples:** Clustering, dimensionality reduction, etc.

# Clustering

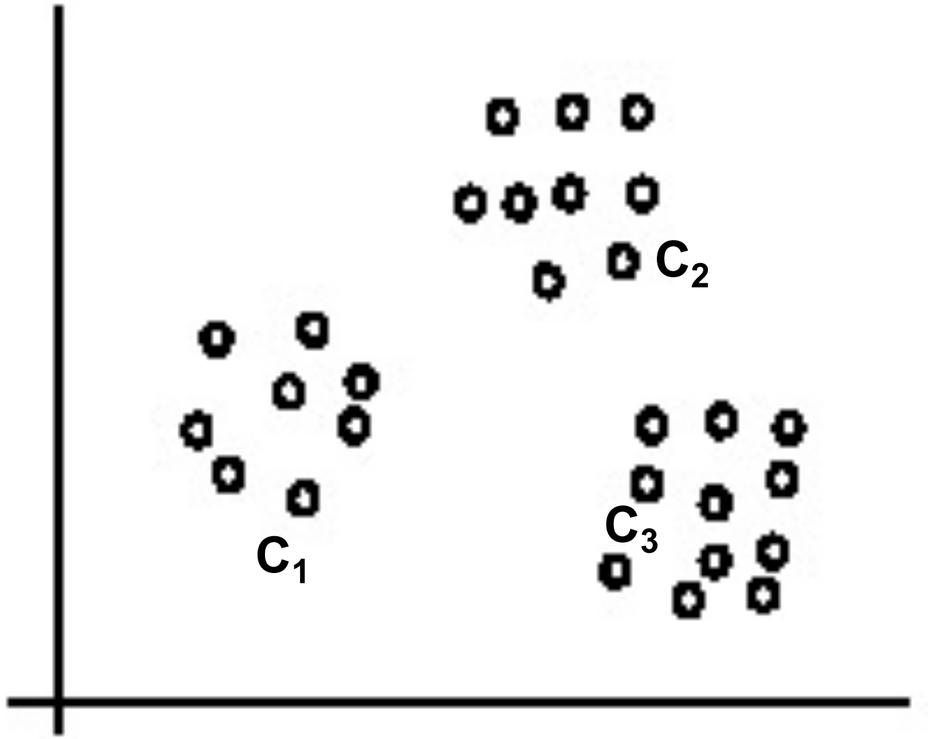
Clustering is a technique for finding similarity groups in data, called clusters. i.e.,

- It groups data instances that are similar to (near) each other in one cluster and data instances that are very different (far away) from each other into different clusters.

Clustering is often called an unsupervised learning task

- No class values

Clustering is often considered synonymous with unsupervised learning.



# Applications Based on Clustering

Example 1: groups people of similar sizes together to make “small”, “medium” and “large” T-Shirts.

- Tailor-made for each person: too expensive
- One-size-fits-all: does not fit all.

Example 2: In marketing, segment customers according to their similarities

- To do targeted marketing.

Example 3: Group a collection of text documents according to their content similarities,

- To produce a topic hierarchy

In fact, clustering is one of the most utilized data mining techniques.

- It has a long history, and used in almost every field, e.g., medicine, psychology, botany, sociology, biology, archeology, marketing, insurance, libraries, etc.

# Aspects of Clustering

A clustering algorithm

- Partitional clustering
- Hierarchical clustering
- …

A distance (similarity, or dissimilarity) function

Clustering quality

- Inter-clusters distance => maximized
- Intra-clusters distance => minimized

The quality of a clustering result depends on the algorithm, the distance function, and the application.


DEEP
LEARNING
INSTITUTE


PRAIRIE VIEW
A&M UNIVERSITY

DLI Accelerated Data Science Teaching Kit

# Thank You