CIS 635 Knowledge Discovery & Data Mining

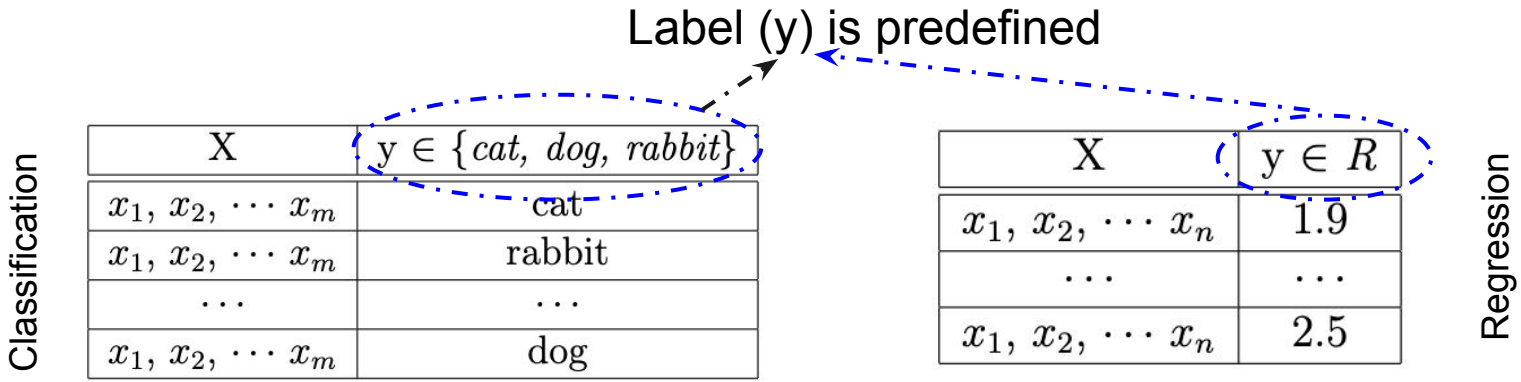
Introduction to Unsupervised Learning

Supervised Learning

- We learned about **Classification** and **Regression**
- These are examples of **supervised** learning

Supervised Learning

- We learned about **Classification** and **Regression**
- These are examples of **supervised** learning
- In your data you have both **X(features)** and **y(Labels)**



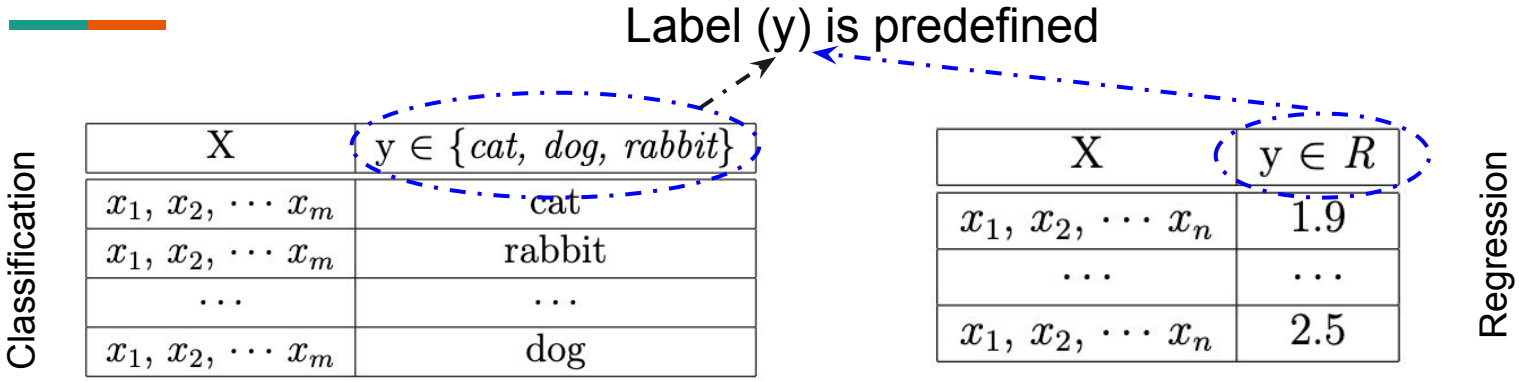
Unsupervised Learning

- In contrast, in **Unsupervised learning**, we have to learn meaningful **representations/models** from **X(features)** only.
- **Clustering** is an example of unsupervised learning

X
$x_1, x_2, \cdots x_o$
$\cdots$
$x_1, x_2, \cdots x_o$

No concept of data label (y)

Unsupervised Learning



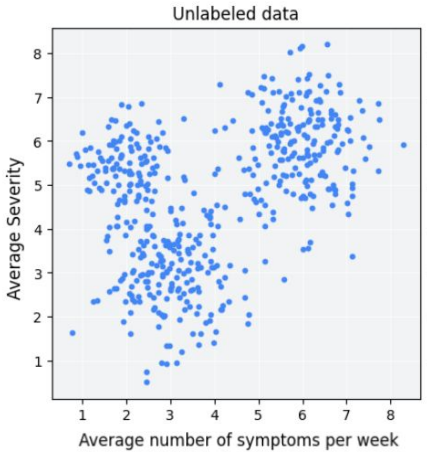
X
$x_1, x_2, \cdots x_o$
$\cdots$
$x_1, x_2, \cdots x_o$

No concept of data label (y)

- Another way to understand the concept of clustering in terms of input data
- For classification/regression we need both **features** and **labels (X, Y)**, where as for clustering we only need **features (X)** as input

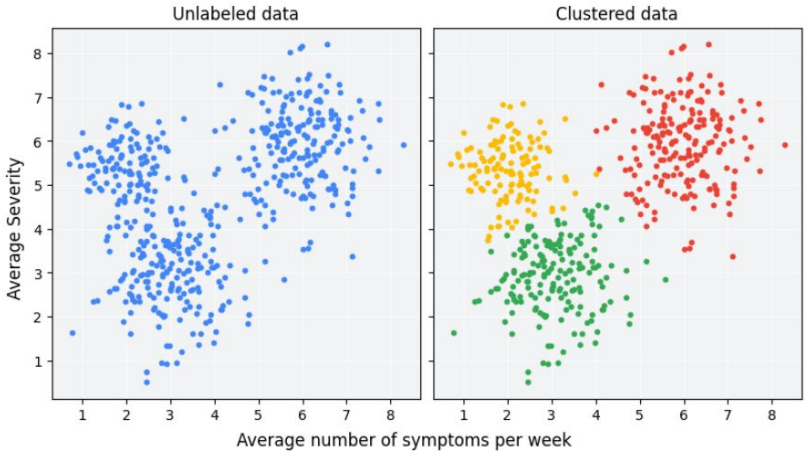
Clustering

- A medical study involving
  - Average number of symptoms per week
  - Average severity



Clustering

- A medical study involving
  - Average number of symptoms per week
  - Average severity



Clustering

- What is clustering?
- **Clustering algorithms:**
  - **K-Means**: Centroid Based
  - **Hierarchical clustering**: Distance connectivity based
  - **GMM**: Distribution based
  - **DBSCAN**: Density Based

- Identifying the number of clusters ?

What is clustering ?

- Grouping of data points based on some features (**X**). Group **indices** are the output

- An example of unsupervised learning

We can use features such as **X {size, color}** to group apples on the right

https://www.codeproject.com/Articles/1273113/Apple-tron-an-AI-for-farmers

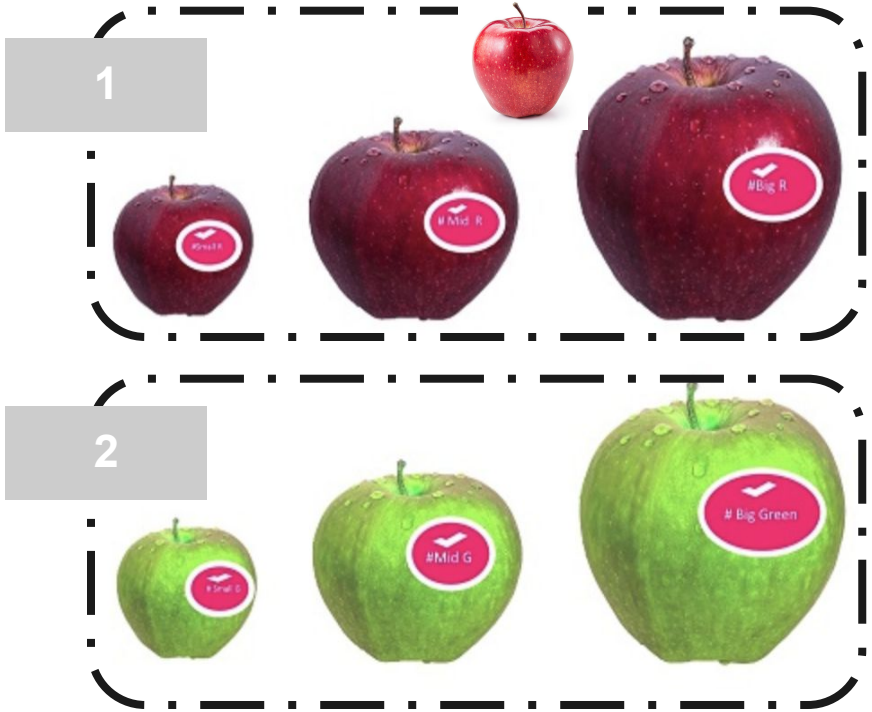
What is clustering ?

- Grouping of data points based on some features (**X**). Group **indices** are the output

- An example of unsupervised learning

- **Labeling a test case:**

  Output: group index **1**



https://www.codeproject.com/Articles/1273113/Apple-tron-an-AI-for-farmers

Based on color

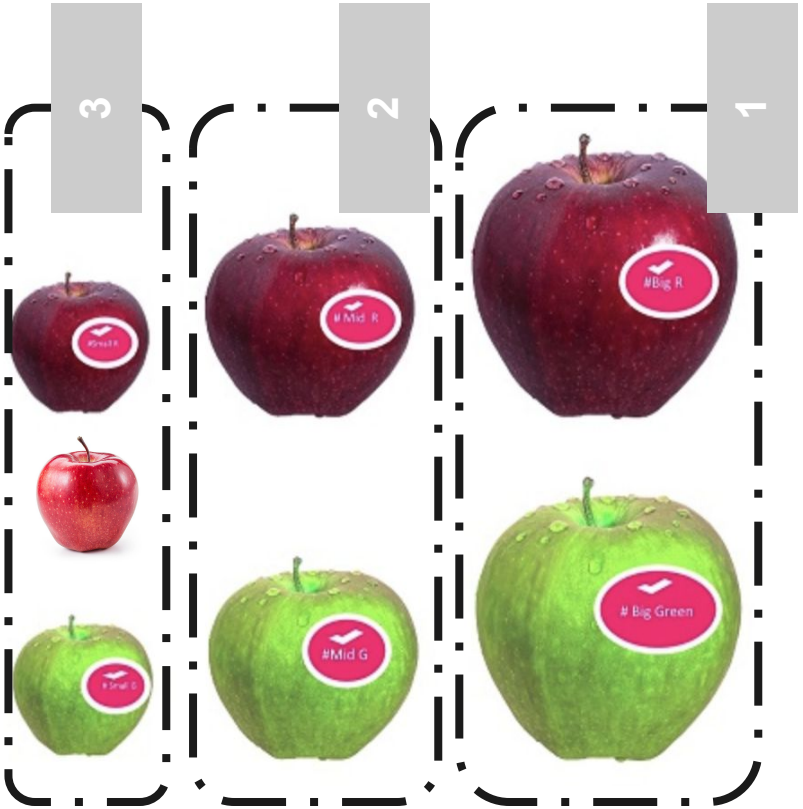
What is clustering ?

- Grouping of data points based on some features (**X**). Group **indices** are the output

- An example of unsupervised learning

- **Labeling a test case:**

  Output: group index **3**



https://www.codeproject.com/Articles/1273113/Apple-tron-an-AI-for-farmers

Based on size

What is clustering ?

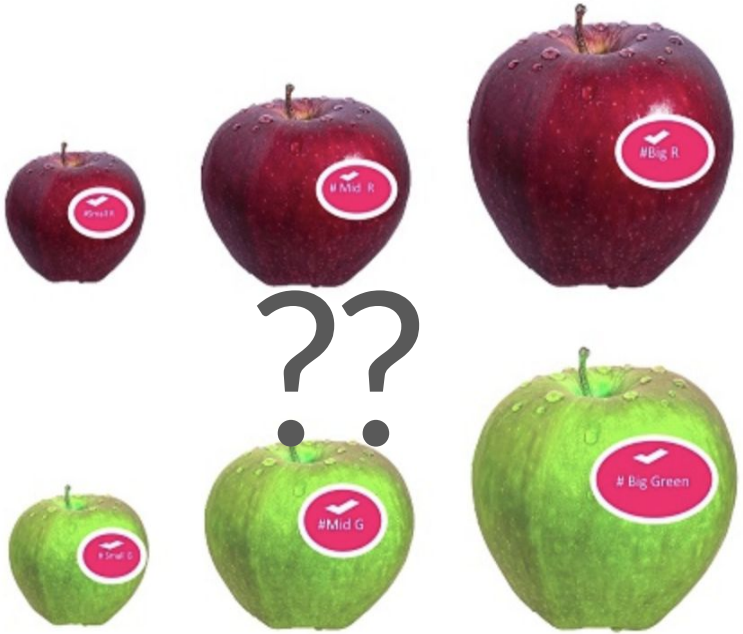
What if we want use both features: size and color?

What is clustering ?

- Grouping of data points based on some features (**X**). Group **indices** are the output

- An example of unsupervised learning

We can use features such as **X {size, color}** to group apples on the right



https://www.codeproject.com/Articles/1273113/Apple-tron-an-AI-for-farmers

Usages of Clustering

Clustering has a myriad of uses in a variety of industries. Some common applications for clustering include the following:

- market segmentation
- social network analysis
- search result grouping
- medical imaging
- image segmentation
- anomaly detection

ref

Usages of Clustering

Clustering has a myriad of uses in a variety of industries. Some common applications for clustering include the following:

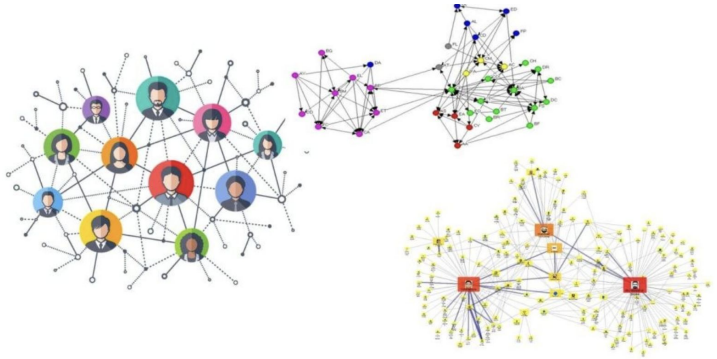
- **Market segmentation**
- Social network analysis
- Search result grouping
- Medical imaging
- Image segmentation
- Anomaly detection

ref

Usages of Clustering

Clustering has a myriad of uses in a variety of industries. Some common applications for clustering include the following:

- Market segmentation
- **Social network analysis**
- Search result grouping
- Medical imaging
- Image segmentation
- Anomaly detection



ref