CIS 678 – Machine Learning

Unsupervised 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)**

Label (y) is predefined

Classification

X	$y \in \{cat,\ dog,\ rabbit\}$
$x_1,\ x_2,\ \cdots x_m$	cat
$x_1,\ x_2,\ \cdots x_m$	rabbit
$\cdots$	$\cdots$
$x_1,\ x_2,\ \cdots x_m$	dog

Regression

| X	$y \in R$
$x_1,\ x_2,\ \cdots x_n$	1.9
$\cdots$	$\cdots$
$x_1,\ x_2,\ \cdots x_n$	2.5

Unsupervised Learning

- In contrast, in **Unsupervised learning**, we have **X(features)**, but we don't have **y(labels)**
- **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

Label (y) is predefined

Classification

X	$y \in \{cat,\ dog,\ rabbit\}$
$x_1,\ x_2,\ \cdots\ x_m$	cat
$x_1,\ x_2,\ \cdots\ x_m$	rabbit
$\cdots$	$\cdots$
$x_1,\ x_2,\ \cdots\ x_m$	dog

X	$y \in R$
$x_1,\ x_2,\ \cdots\ x_n$	1.9
$\cdots$	$\cdots$
$x_1,\ x_2,\ \cdots\ x_n$	2.5

Regression

X
$x_1,\ x_2,\ \cdots\ x_o$
$\cdots$
$x_1,\ x_2,\ \cdots\ x_o$
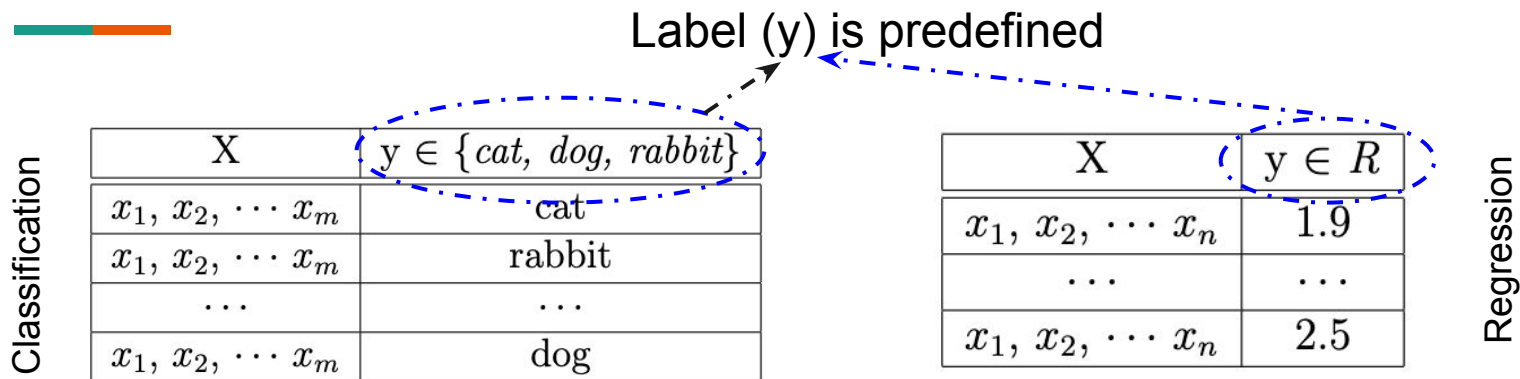
No concept of data label (y)

- 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

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

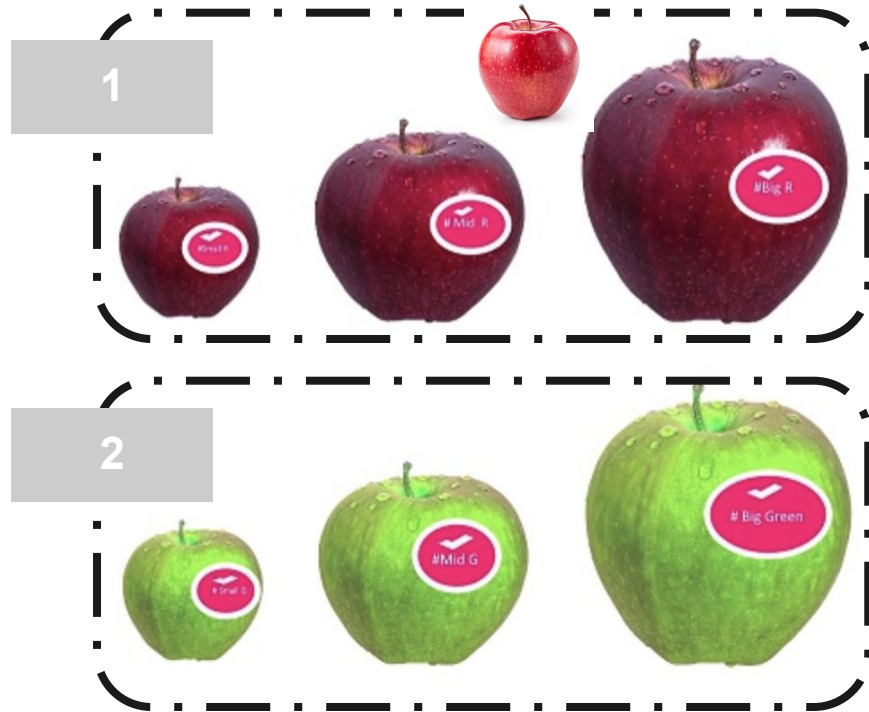- Identifying the number of clusters ?

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

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:**
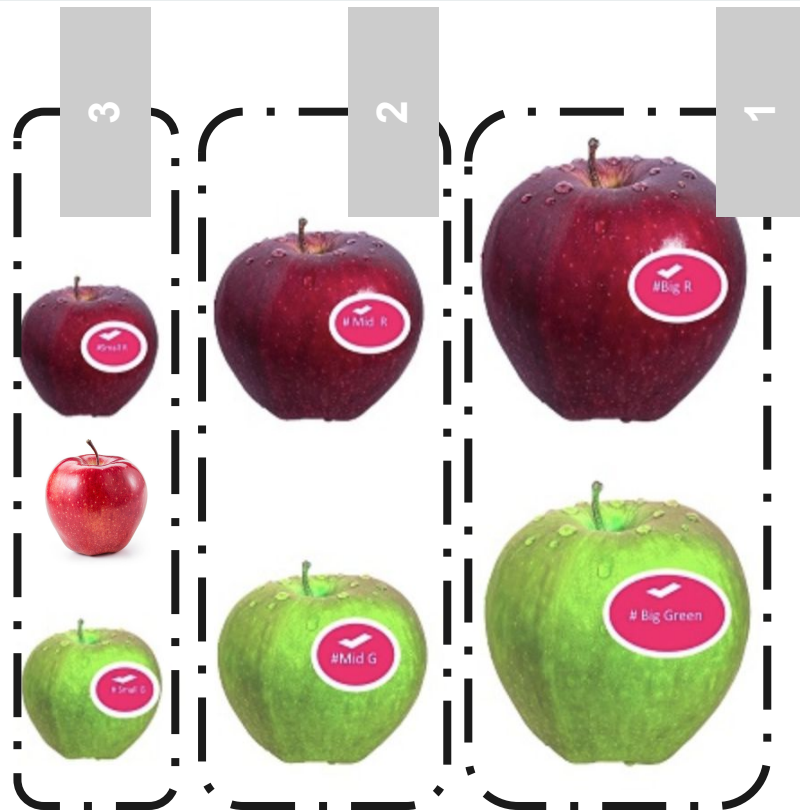
    Output: group index **1**

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