



# CIS 678 - Machine Learning

Introduction to Unsupervised Learning



# Supervised Learning

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



## Supervised (Discriminative) Learning

$$f(y='diabetes'|X=['age', 'bmi'])$$



## Supervised (Discriminative) Learning

$$f(y='diabetes'|X=['age', 'bmi'])$$

$$f(y|X)$$

*Essentially we are learning the relationship  
between  $X$  and  $y$*

# Supervised (Discriminative) Learning

$$y \in R$$

- Insurance cost :  $y \in R$
- House price:  $y \in R$
- Weather prediction:  $Y \in R$
- Energy consumption:  $y \in R$
- Sales forecasting:  $y \in R$

Regression

$$y \in \{....\}$$

- Cat vs dog:  $y \in \{\text{cat}, \text{dog}\}$
- Spam filter:  $y \in \{\text{spam}, \text{not-spam}\}$
- Digit classification:  $y \in \{0, 1, ..., 9\}$
- Sentiment classification:  $y \in \{\text{happy}, \text{sad}, \text{confused}, \text{angry} \dots\}$

*note: has to be from a closed set*

Classification

# 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	$y \in \{cat, dog, rabbit\}$
$x_1, x_2, \dots x_m$	cat
$x_1, x_2, \dots x_m$	rabbit
...	...
$x_1, x_2, \dots x_m$	dog

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

Regression



# 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
$x_1, x_2, \dots x_o$
$\dots$
$x_1, x_2, \dots x_o$

No  
concept  
of data  
label (y)



# 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

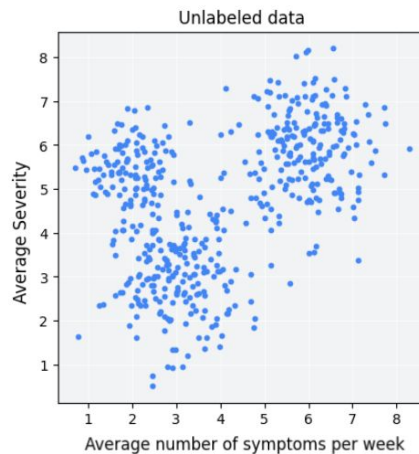
X	No concept of data label (y)
$x_1, x_2, \dots x_o$	
$\dots$	
$x_1, x_2, \dots x_o$	

$f(X)$



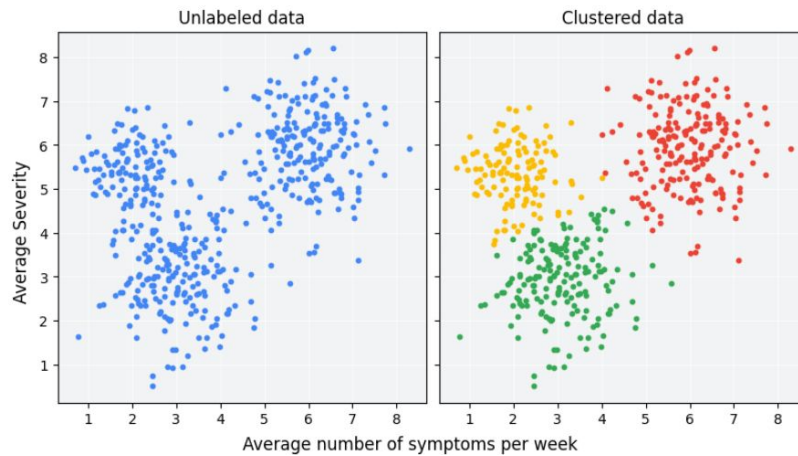
# Clustering

- 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

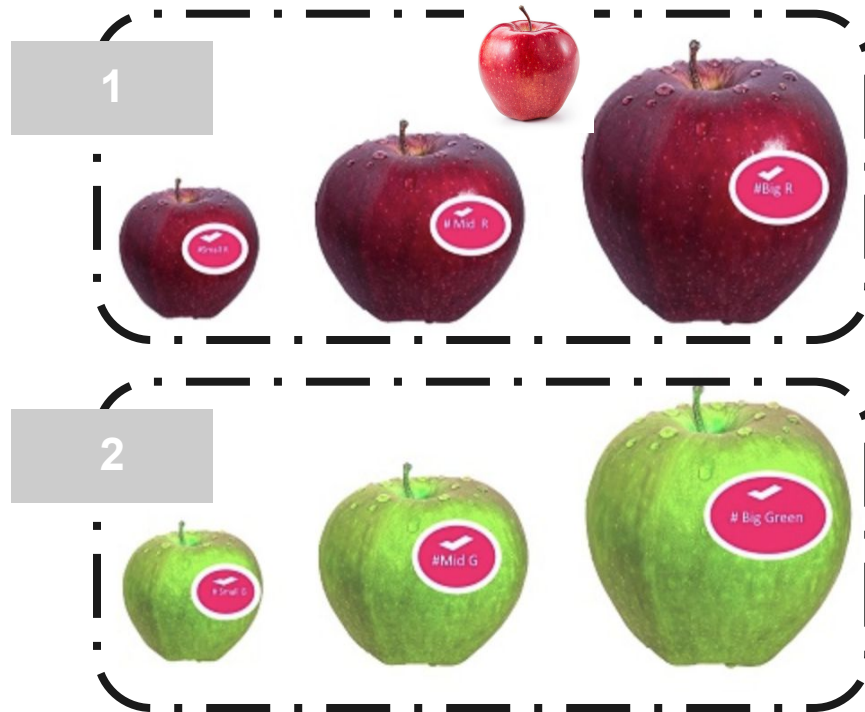
We can use features such as **X {size, color}** to group apples on the right



# 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

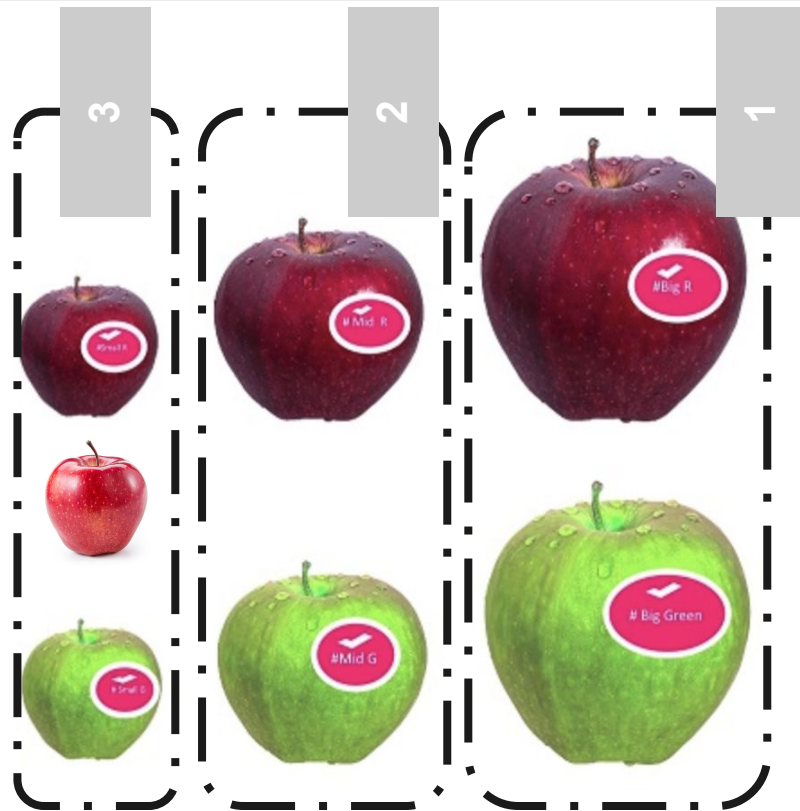


<https://www.codeproject.com/Articles/1273113/Apple-tron-an-AI-for-farmers>

**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 ?

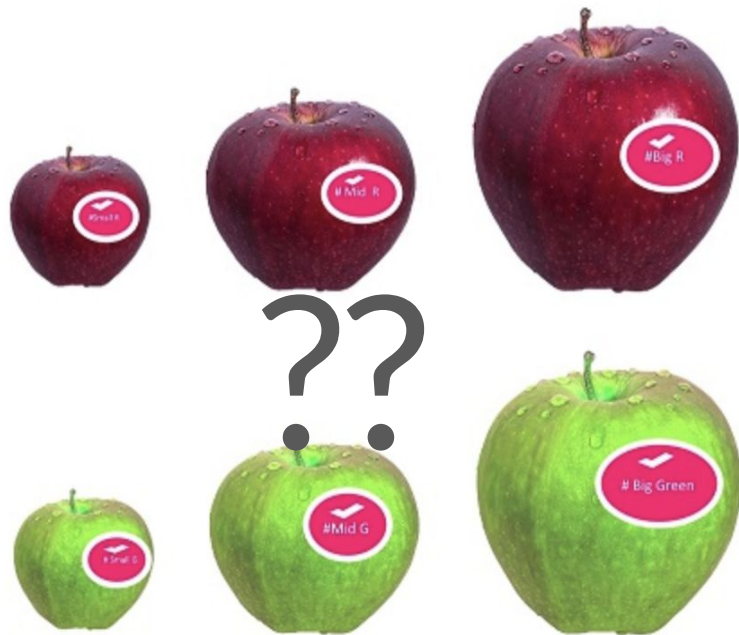


What if we want use both features: size and color?

# 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







# 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

# Usages of Clustering

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

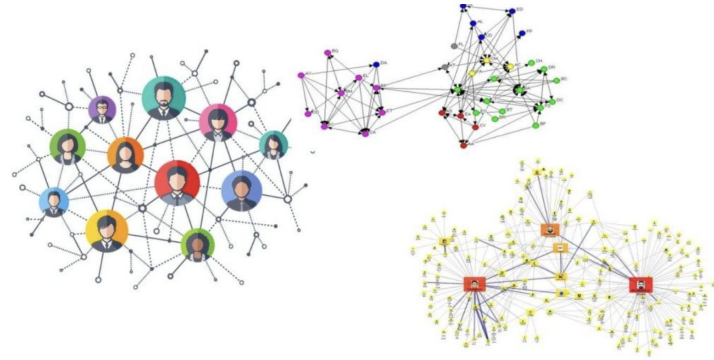
- **Market segmentation**
- Social network analysis
- Search result grouping
- Medical imaging
- Image segmentation
- Anomaly detection



# 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





**QA**