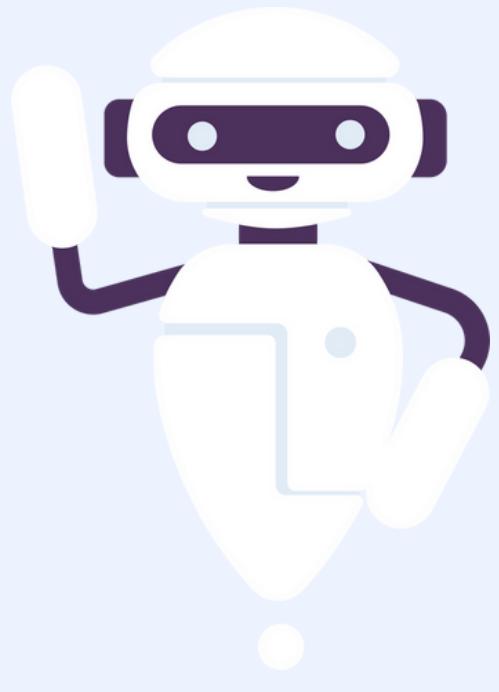
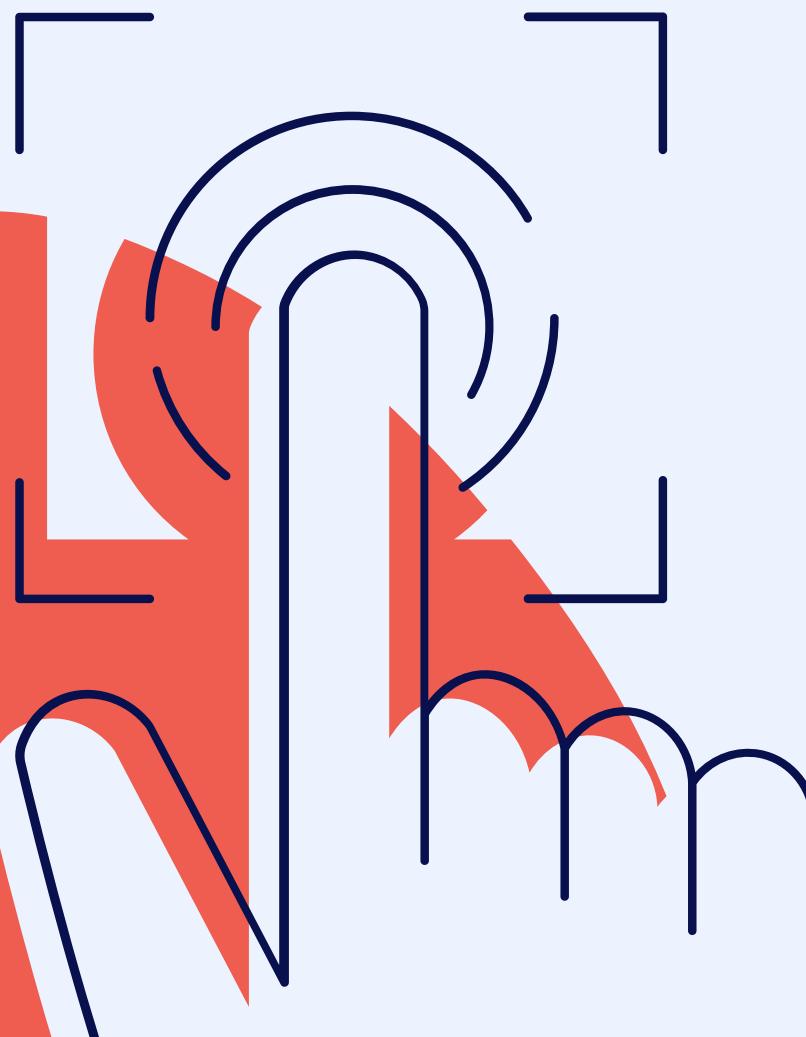




# Beyond Labels: Exploring the Potential of Unsupervised Learning

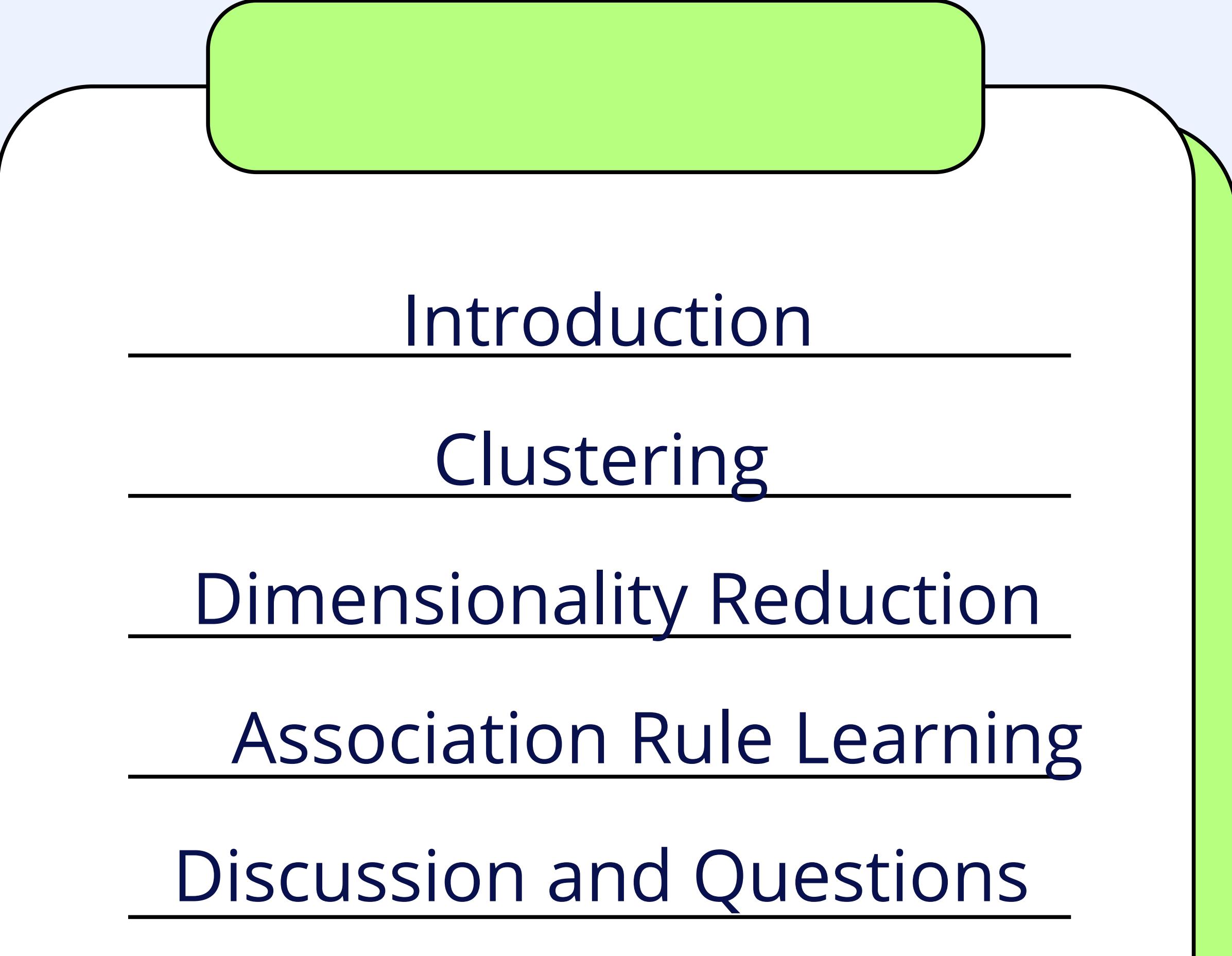


# About me

- **4th Year Student at "ESI  
Alger" Computer Systems  
Option.**
- **Technical Department Co-  
manager at School of AI  
Algiers.**
- **I Love Data Science <3 !**



# What's on the menu ?



Introduction

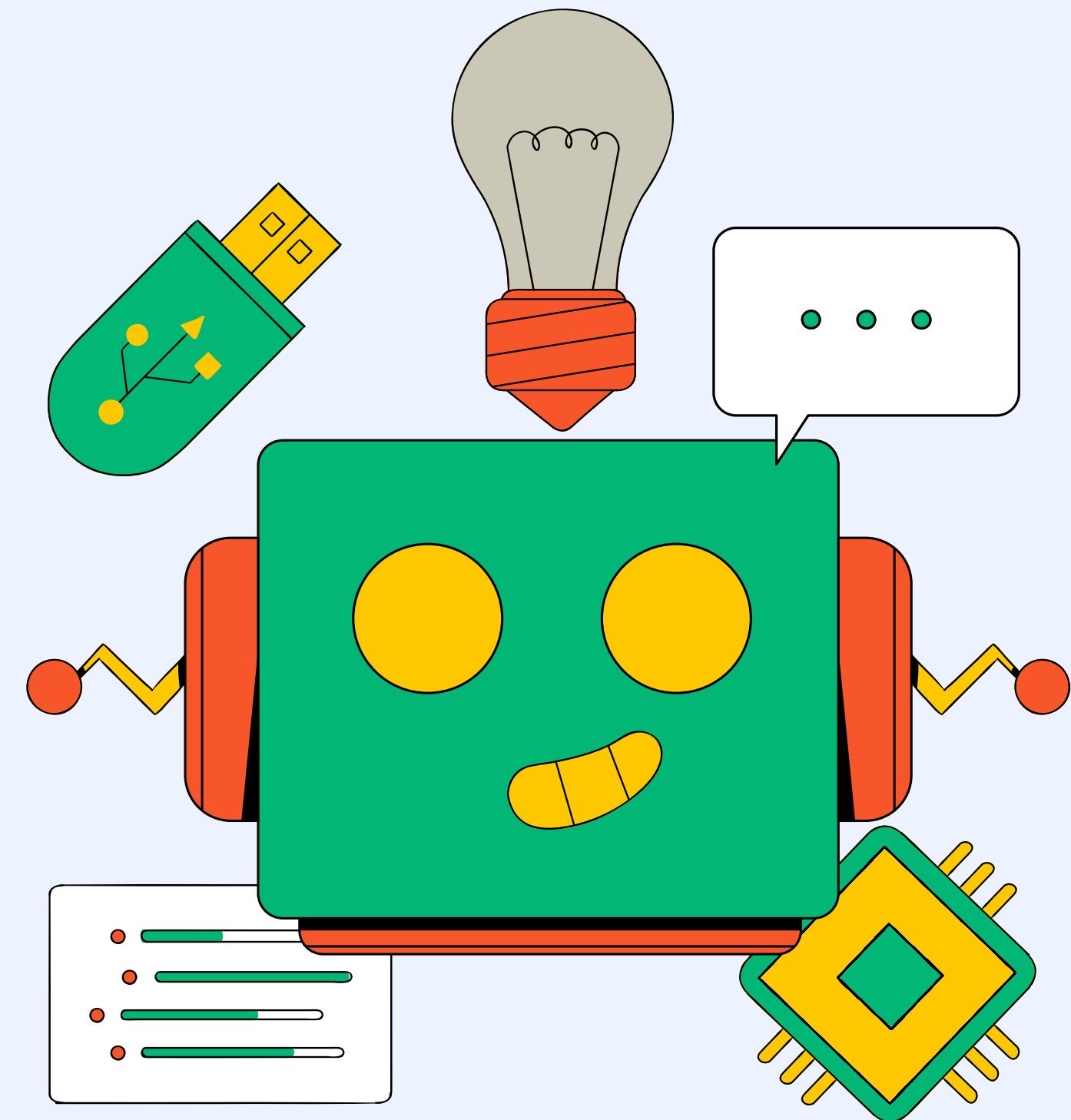
Clustering

Dimensionality Reduction

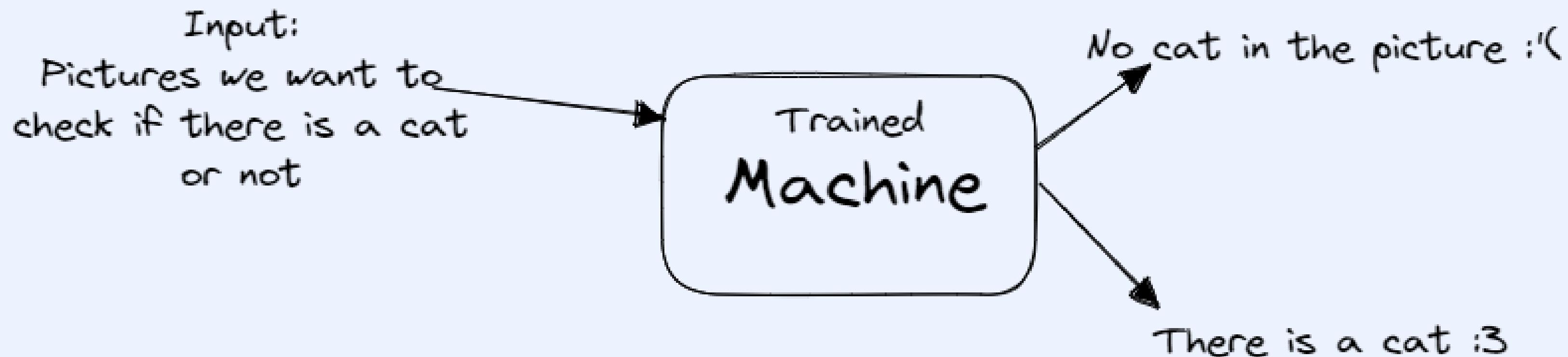
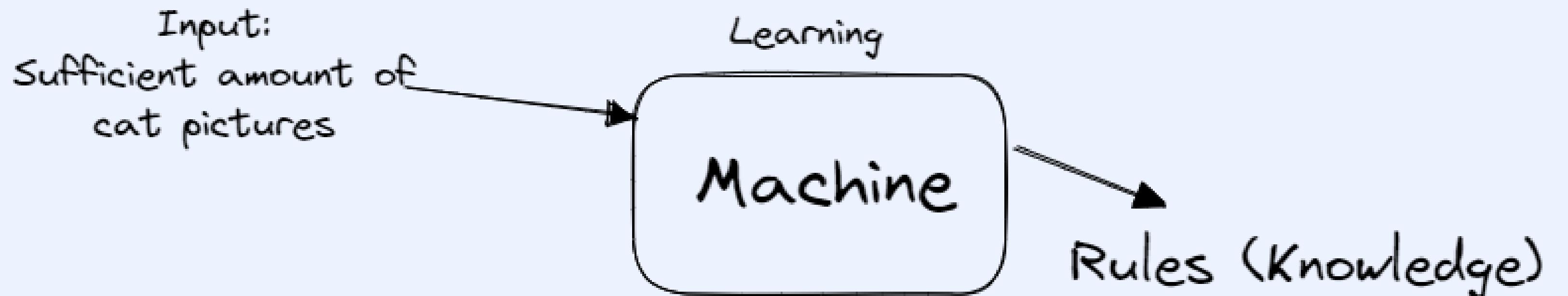
Association Rule Learning

Discussion and Questions

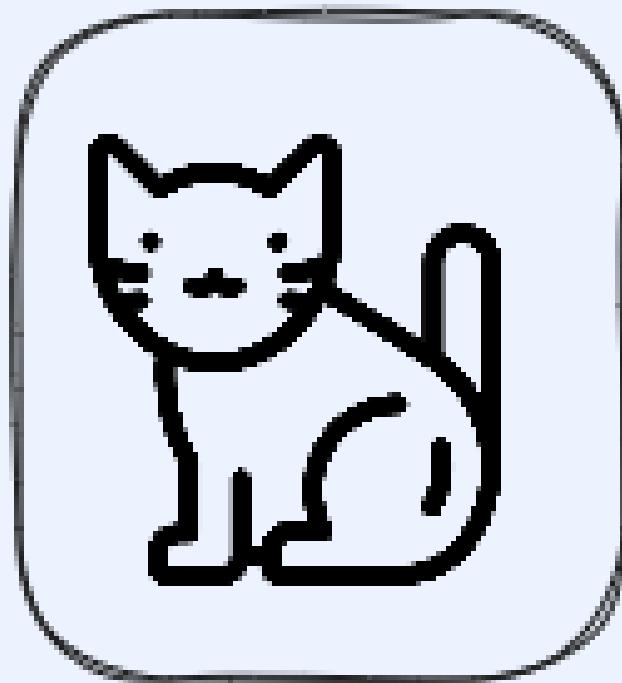
# Introduction



# Machine Learning



# Supervised Learning

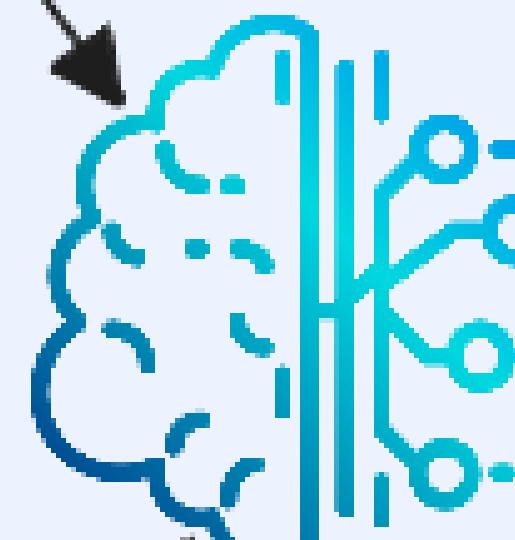


Data

Those are Cats

Labels

Training the model  
to detect cats



I can differentiate between cats  
and anything else

# Rule Time !

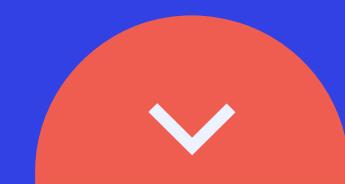
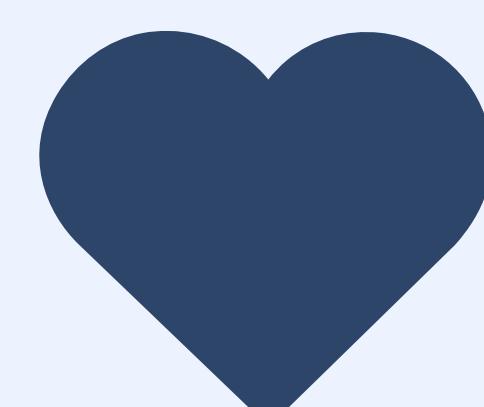
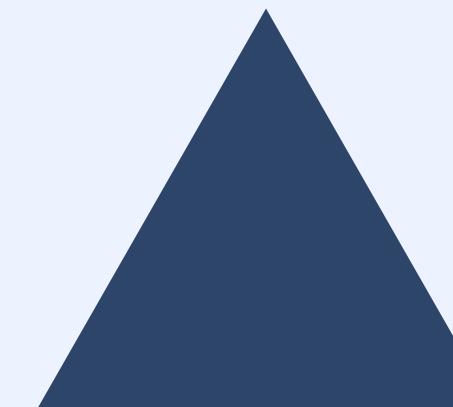
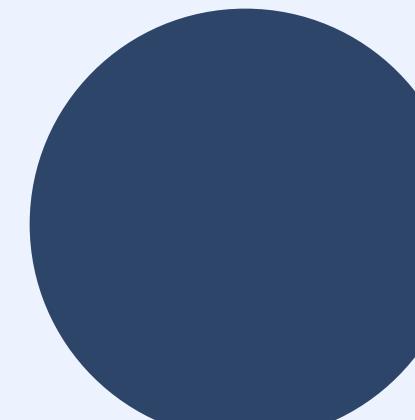
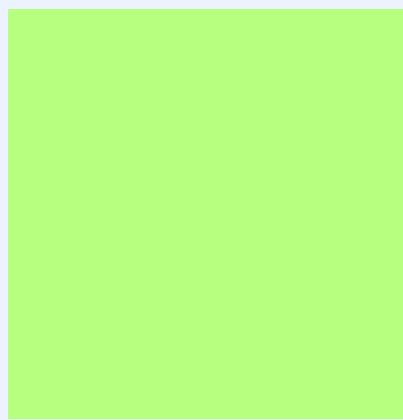
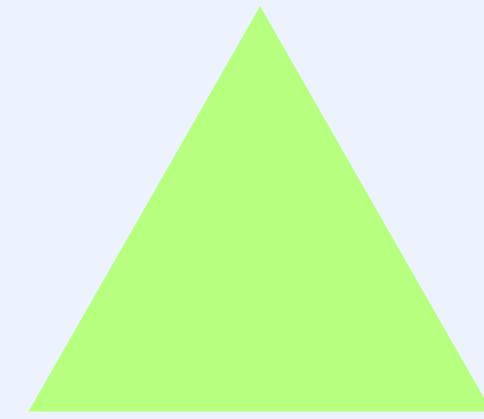
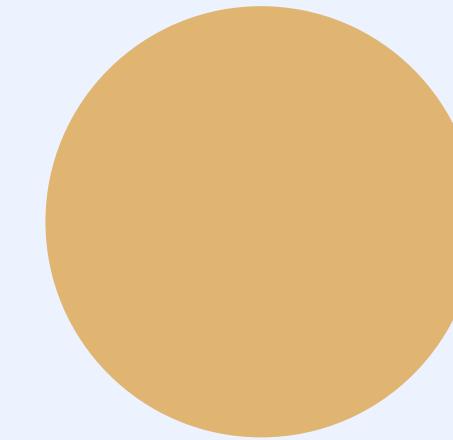
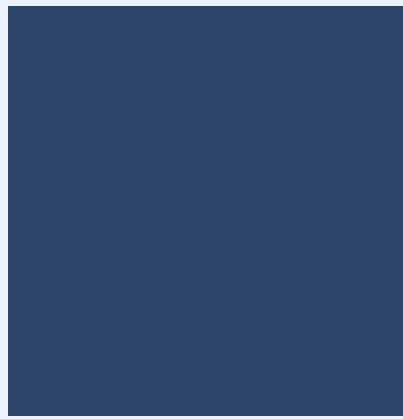
## SUPERVISED LEARNING

---

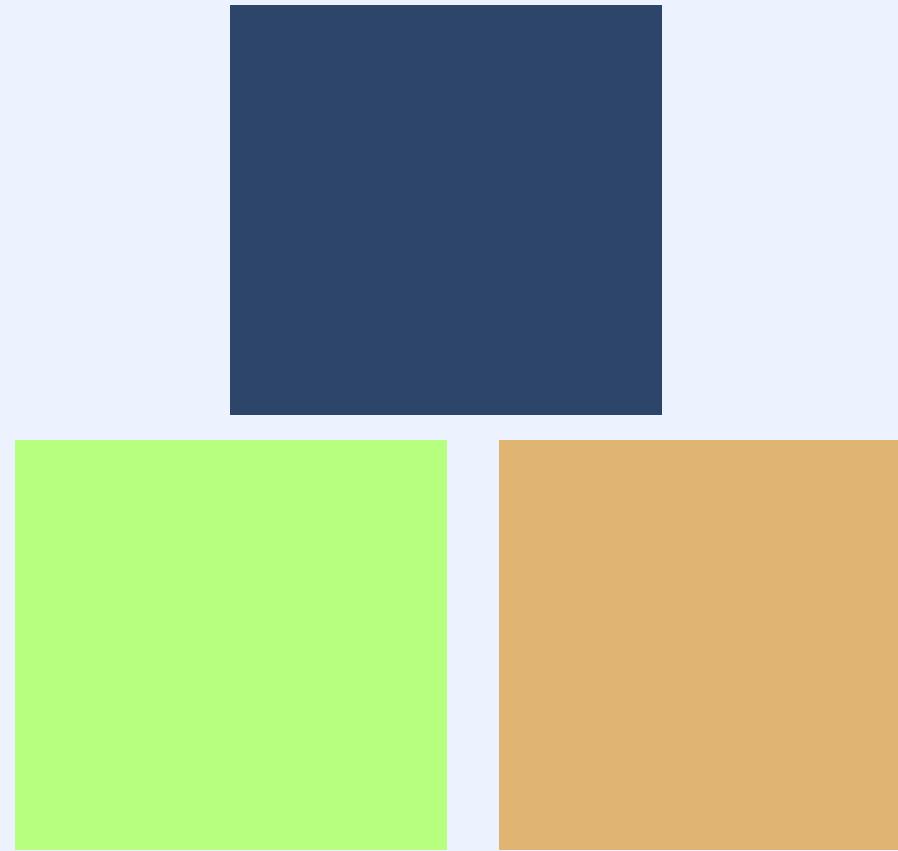
In supervised learning, we're trying to build a model to **predict** an **answer** or **label** provided by a teacher.



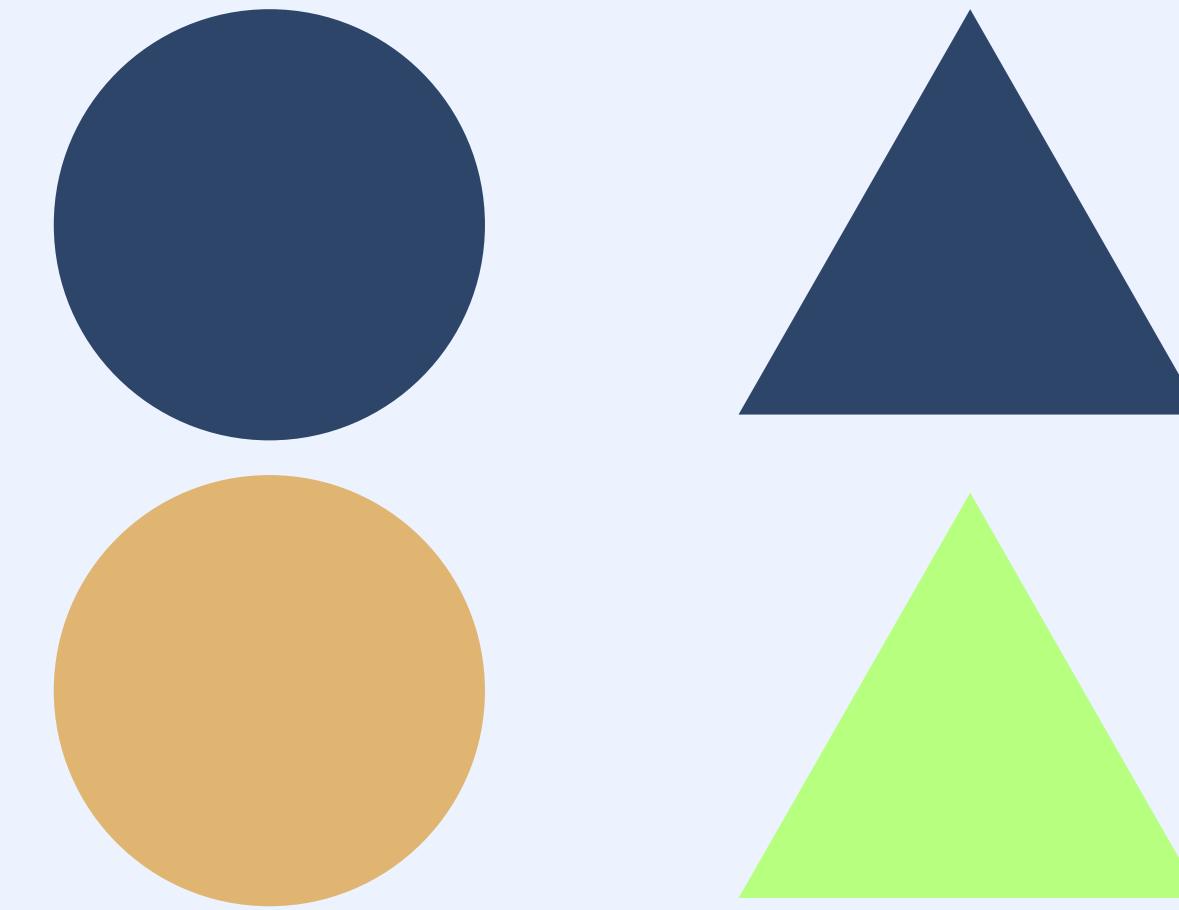
# Unsupervised Learning



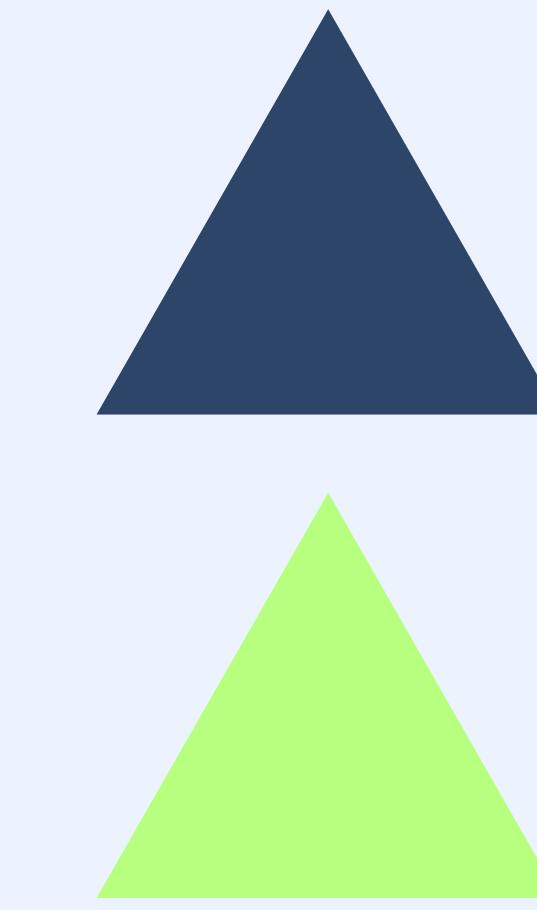
# Unsupervised Learning



**Squares**



**Circles**



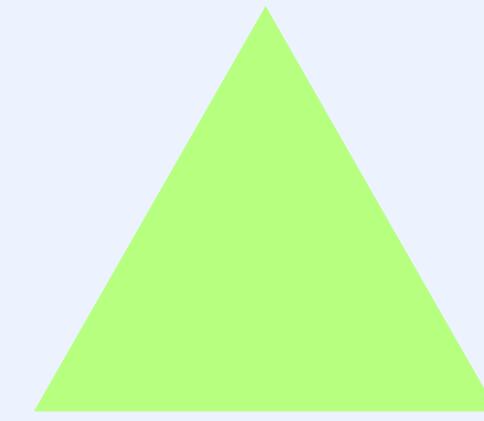
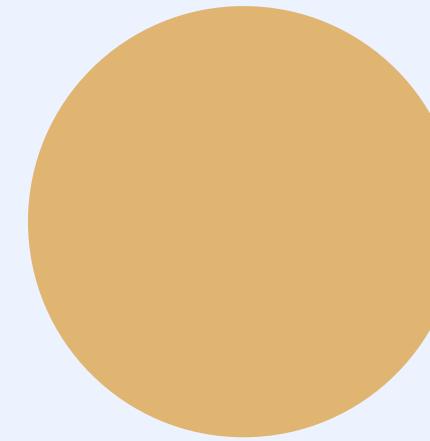
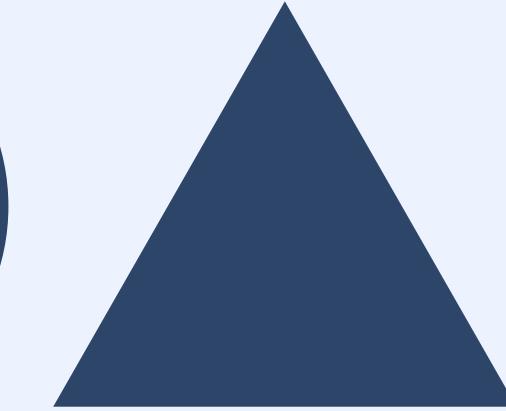
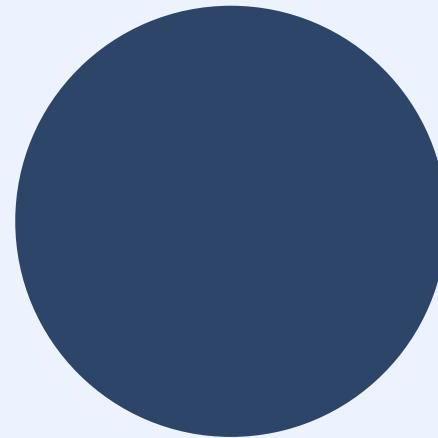
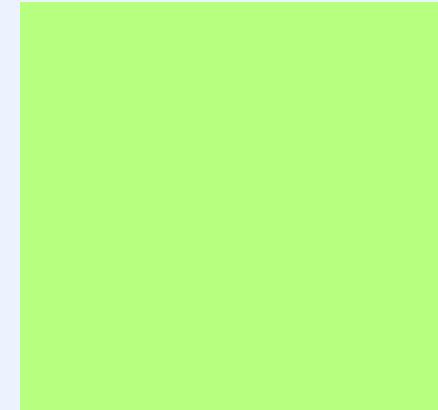
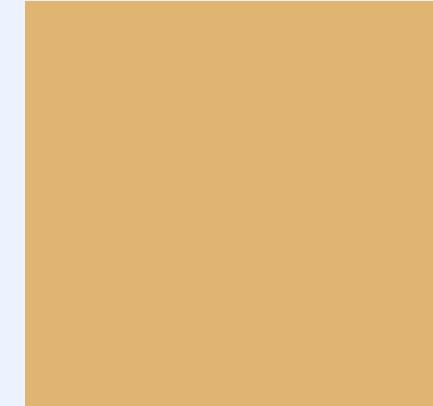
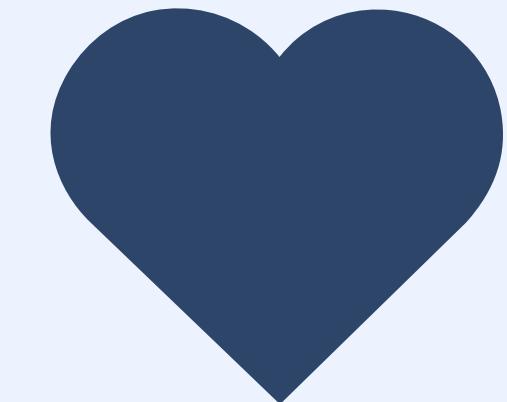
**Triangles**



**Heart**



# Unsupervised Learning



Blue

Yellow

Green



# Rule Time !

## UNSUPERVISED LEARNING



In unsupervised learning, instead of a teacher, the world around us is basically providing training labels.



# Most Performant Unsupervised Learning Models:



# Supervised vs Unsupervised •



Labels !



Goals !



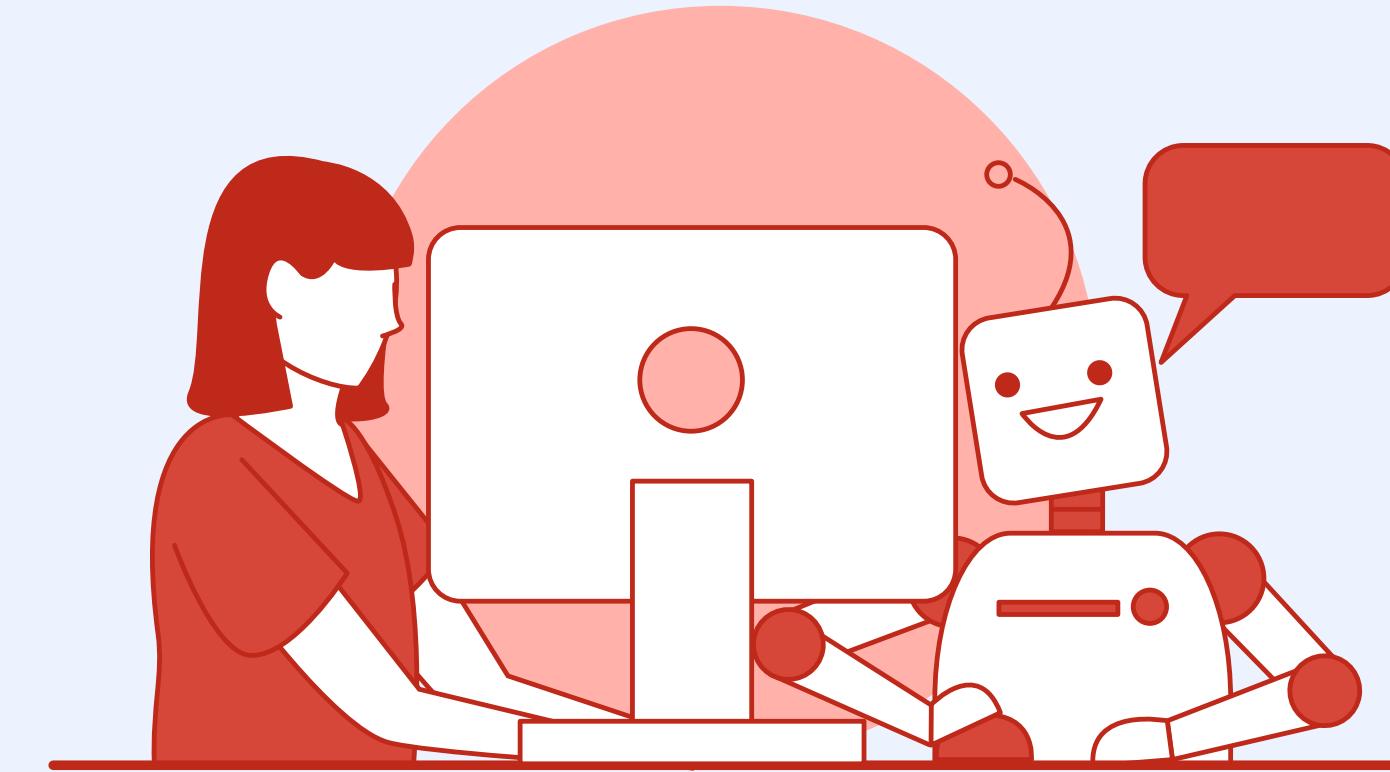
Applications !



Complexity !

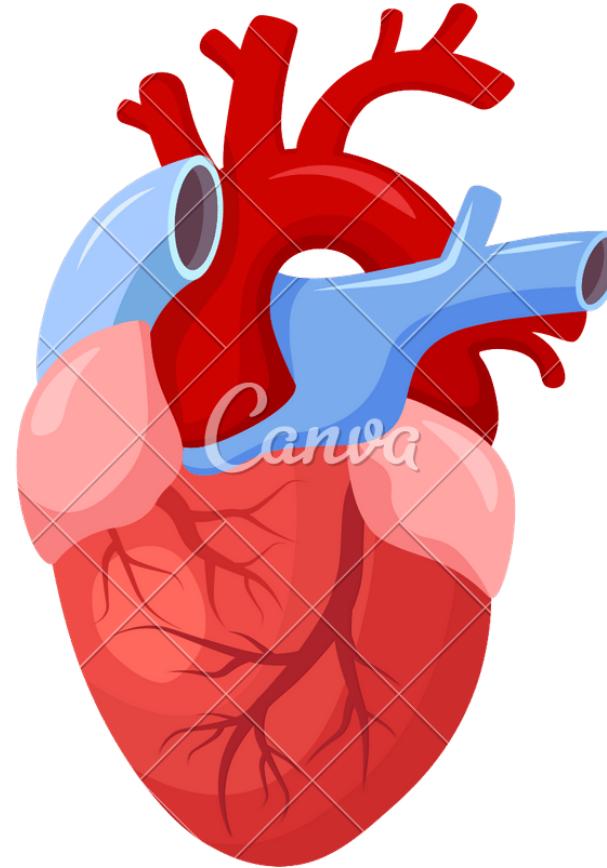


Drawbacks !



# Real-World Examples:

o



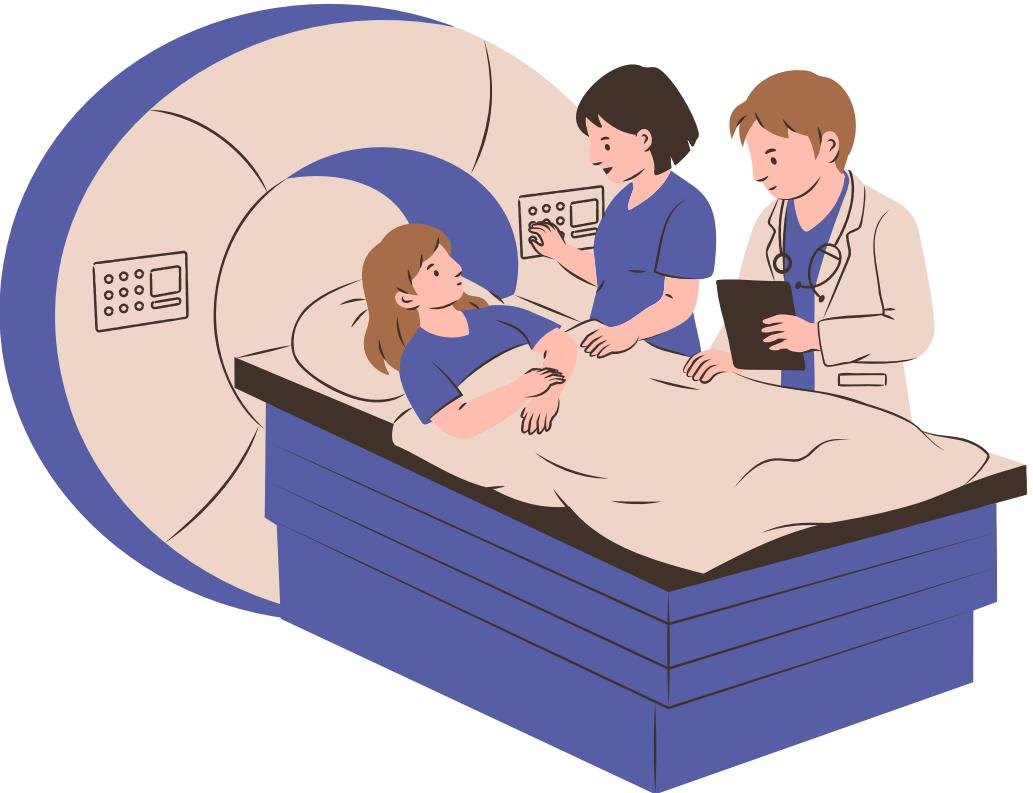
**Heart  
Disease  
Detection**



**Customer  
Segmentation  
in Marketing**



**Content  
Recommendation**

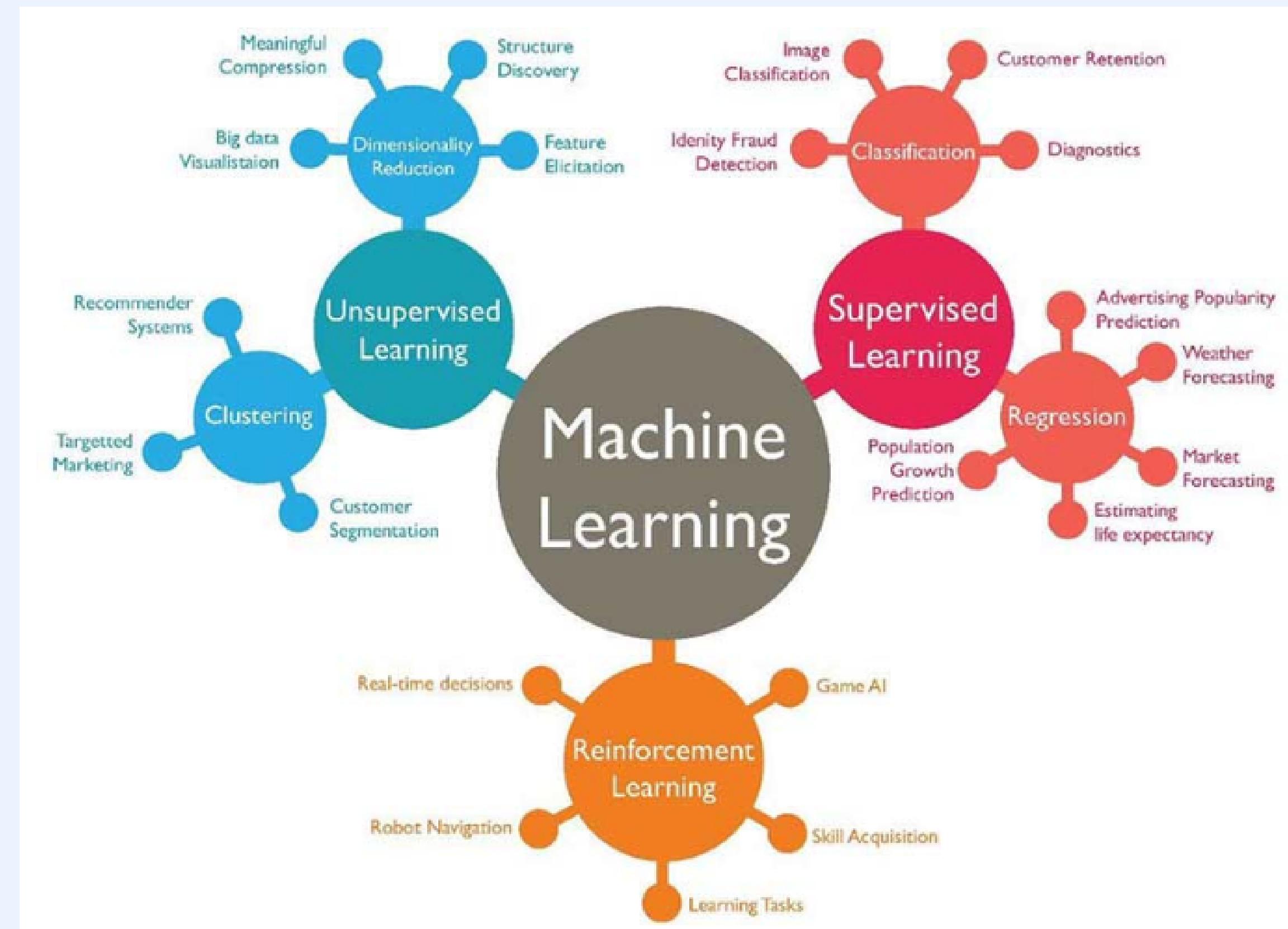


**Medical Imaging**

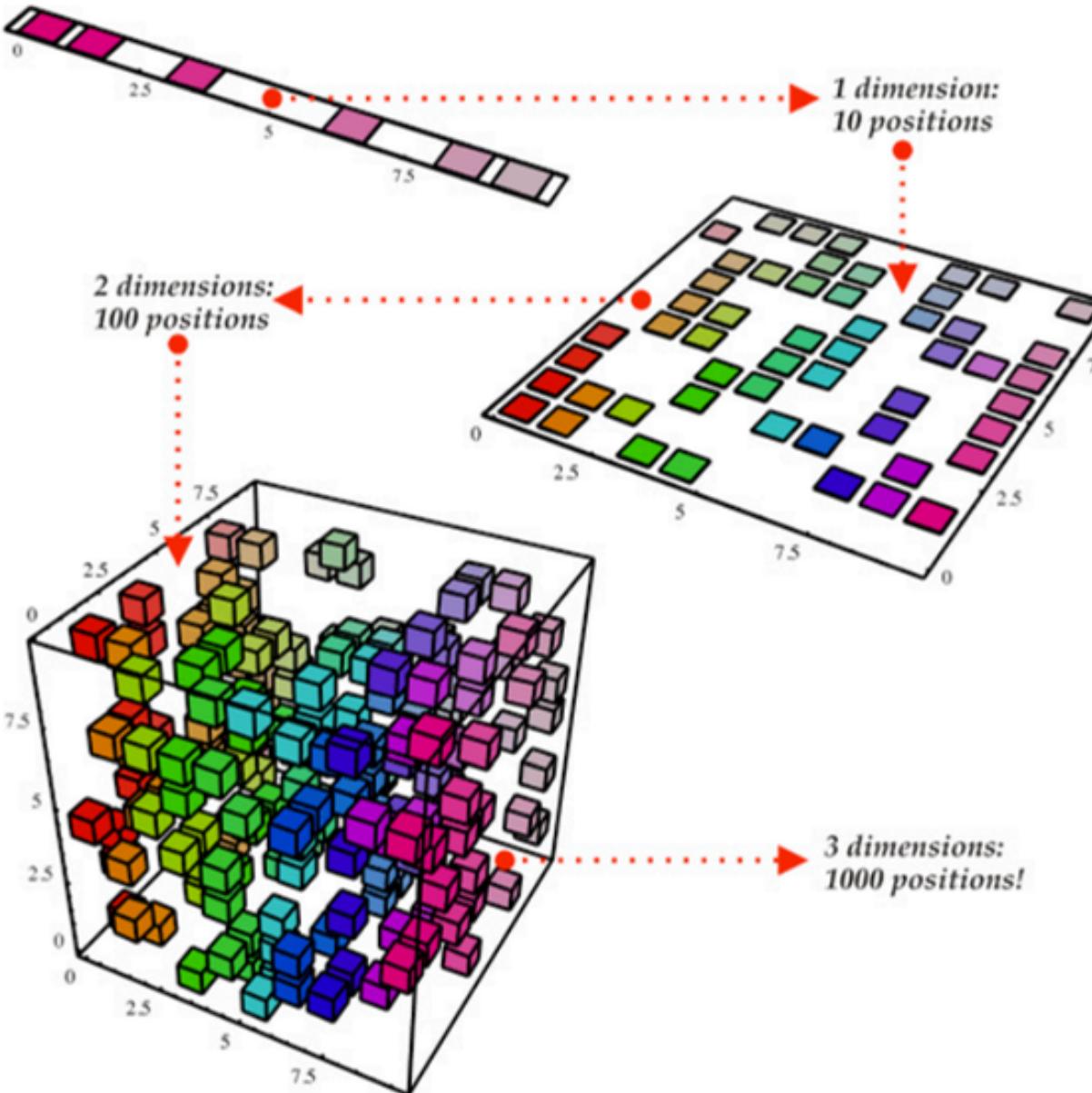


x

# Types of Learning

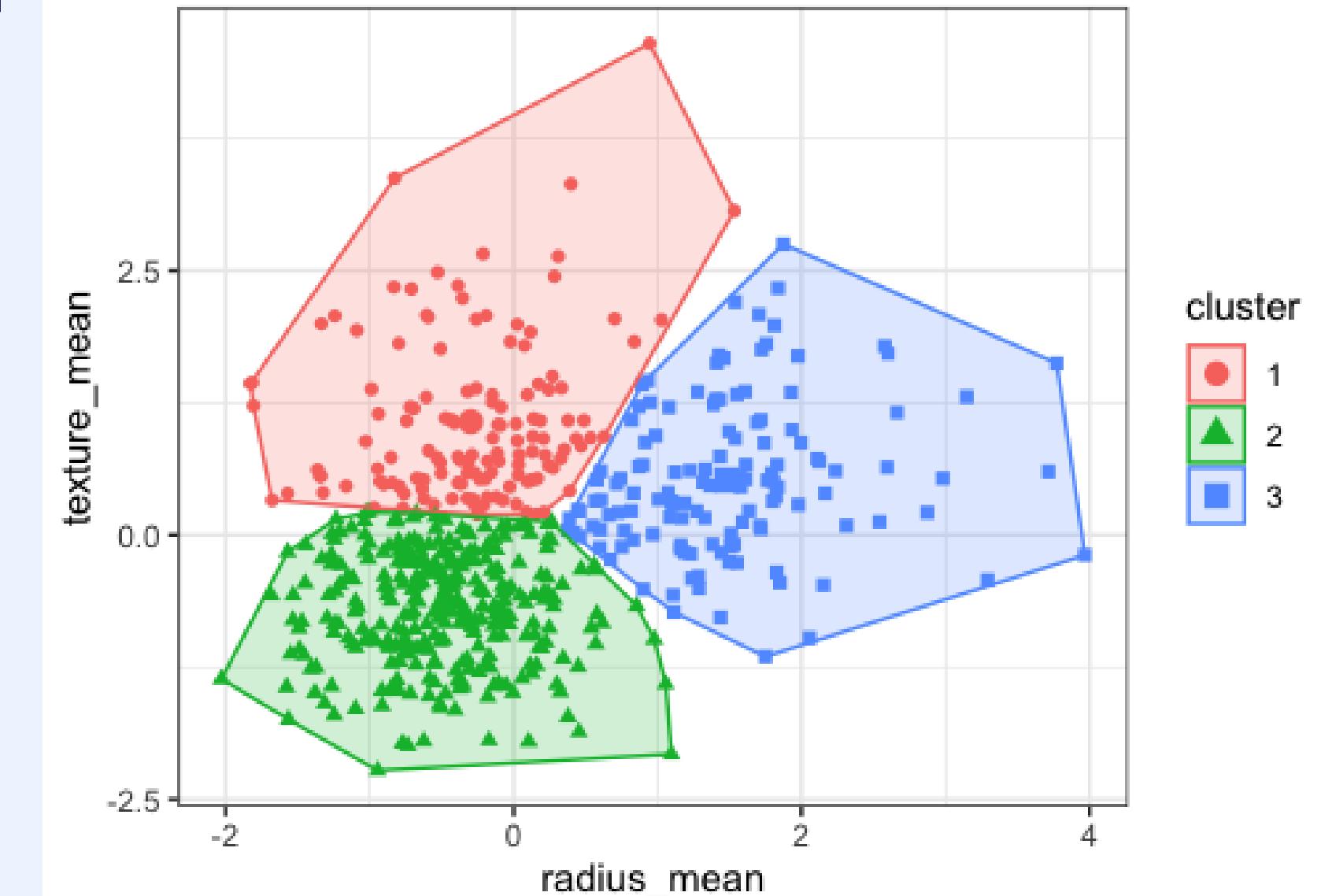


# Types of Unsupervised Learning



Dimensionality Reduction

Cluster plot



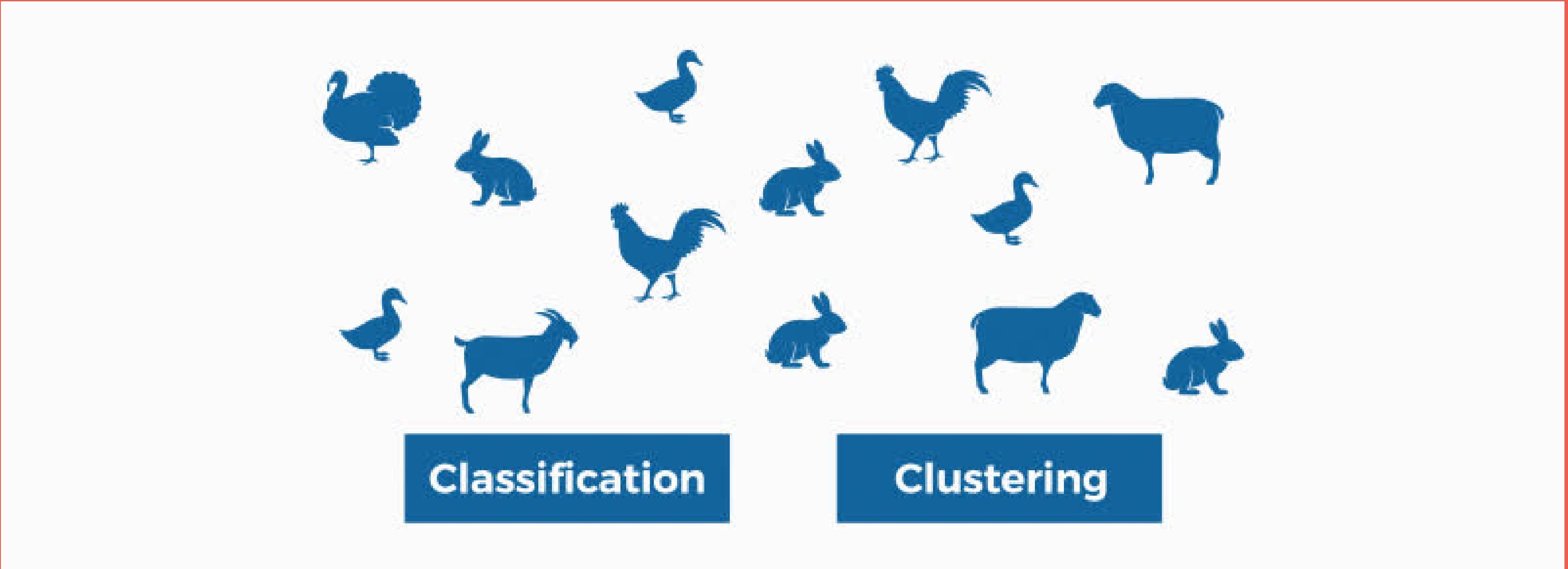
Clustering

# Guess the type !

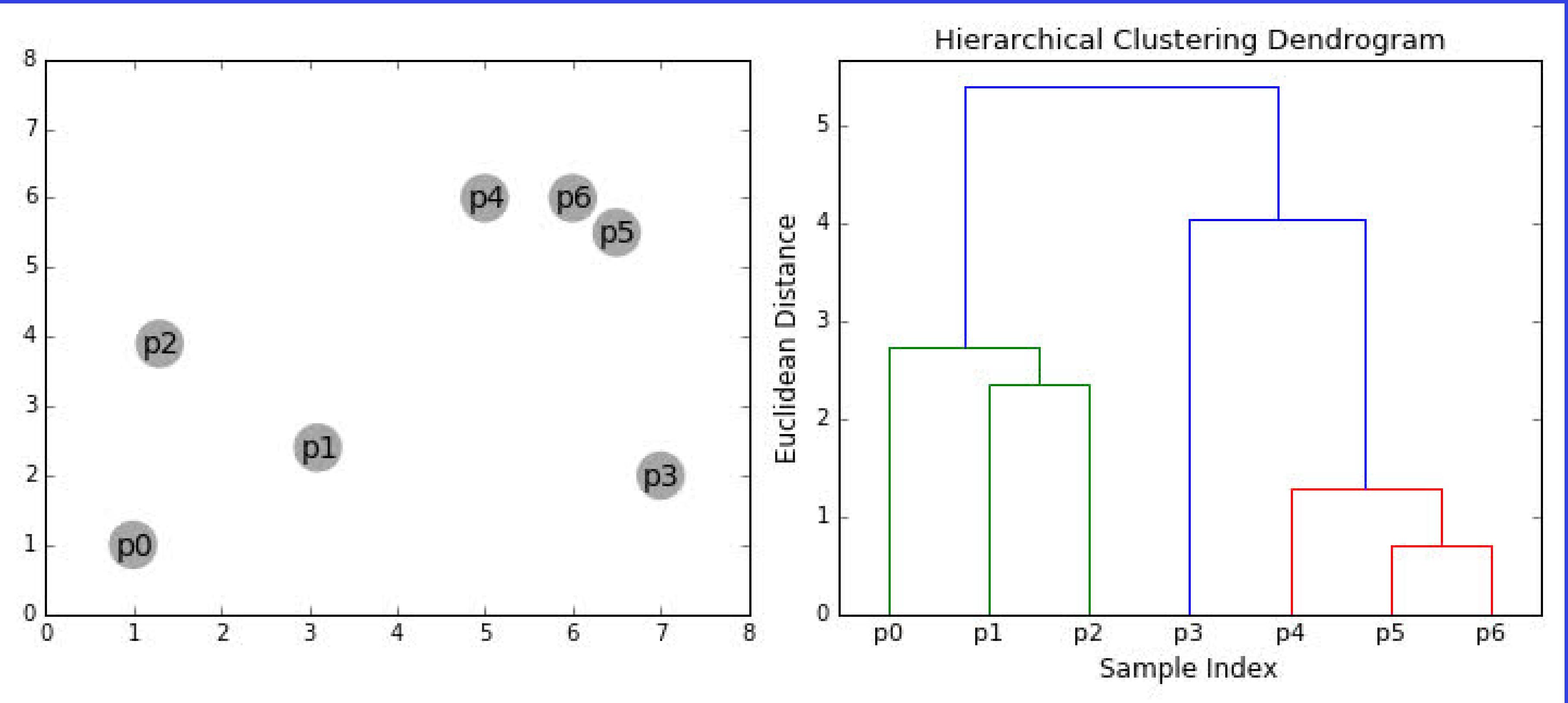
1. Predict whether a user will make a purchase in the next session using historical previous data.
2. Categorize songs into different genres using beats per minute, song duration, and audio level.
3. Recommend friends in Social Medias: we have data about who is connected with whom, but we don't have any specific criteria for what makes two people likely to be friends
4. Automatically group news articles into different topics using words and phrases similarity.

# Clustering

x



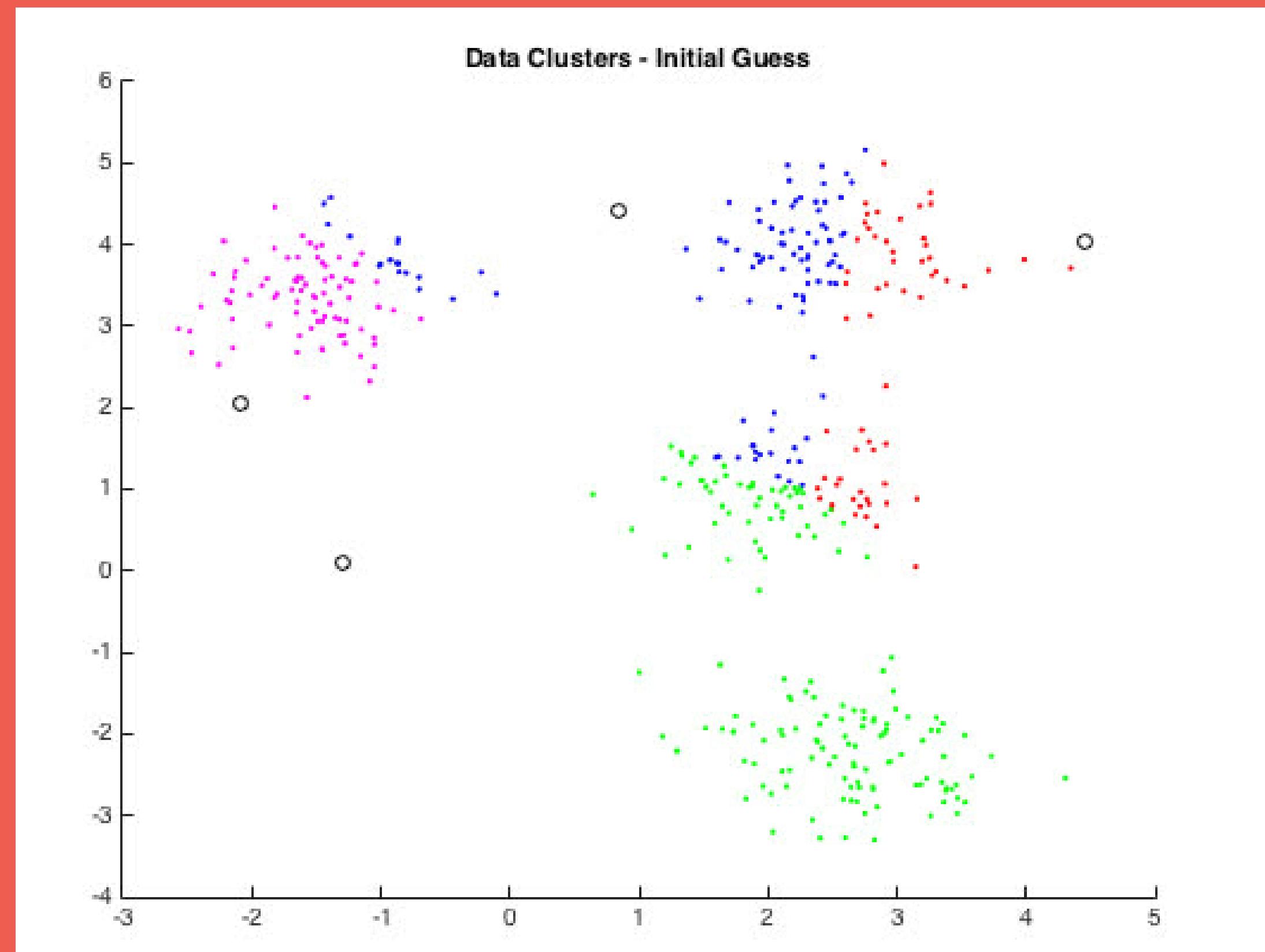
# Hierarchical Clustering



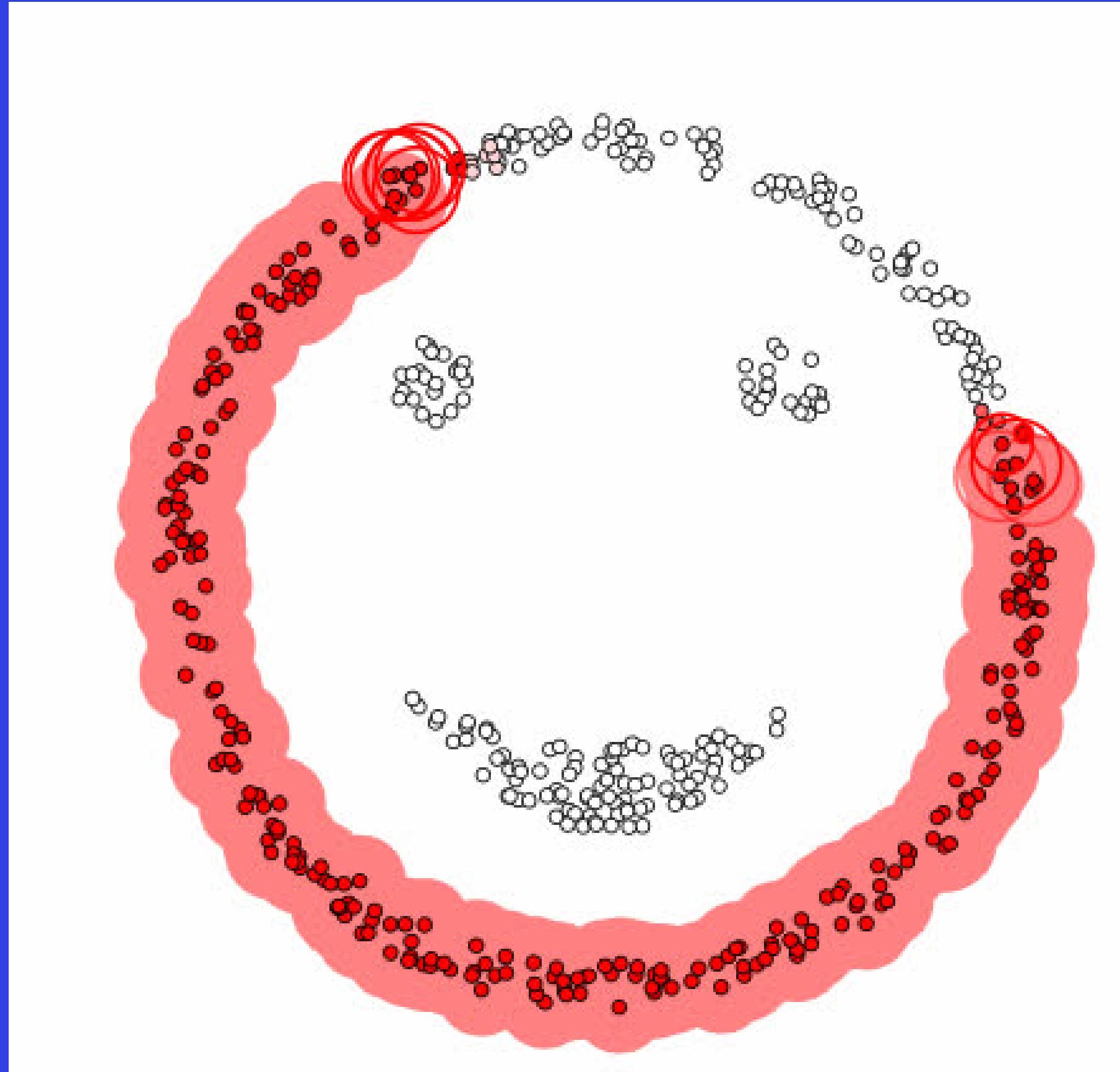
# K-Means

X

O



# DBSCAN



# QUESTION !

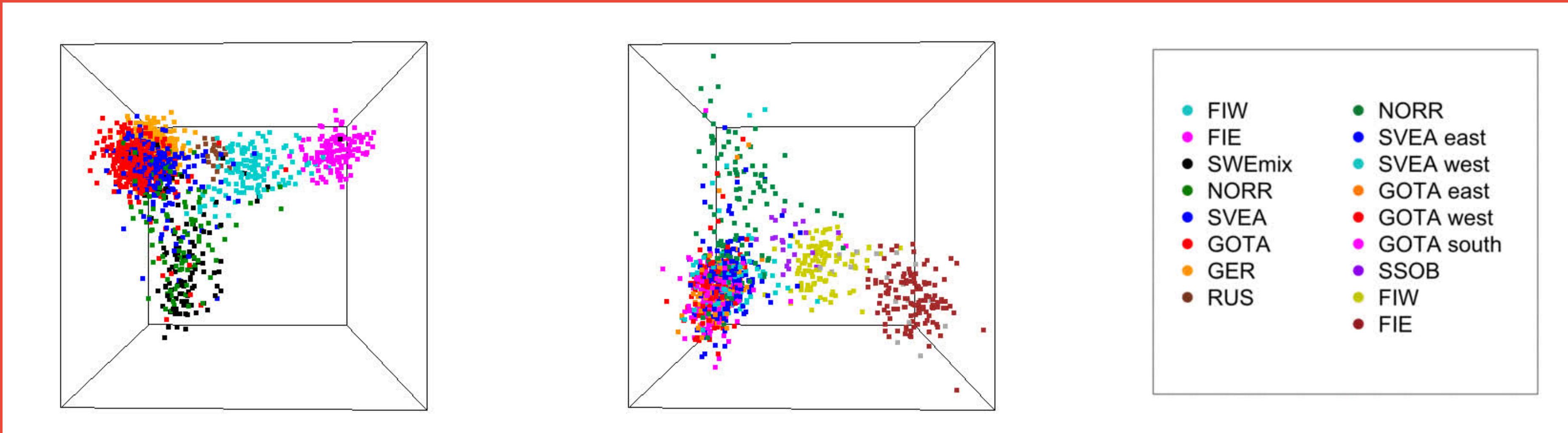
How do hierarchical and density-based clustering algorithms handle noisy data compared to K-means?



# PCA

X

O

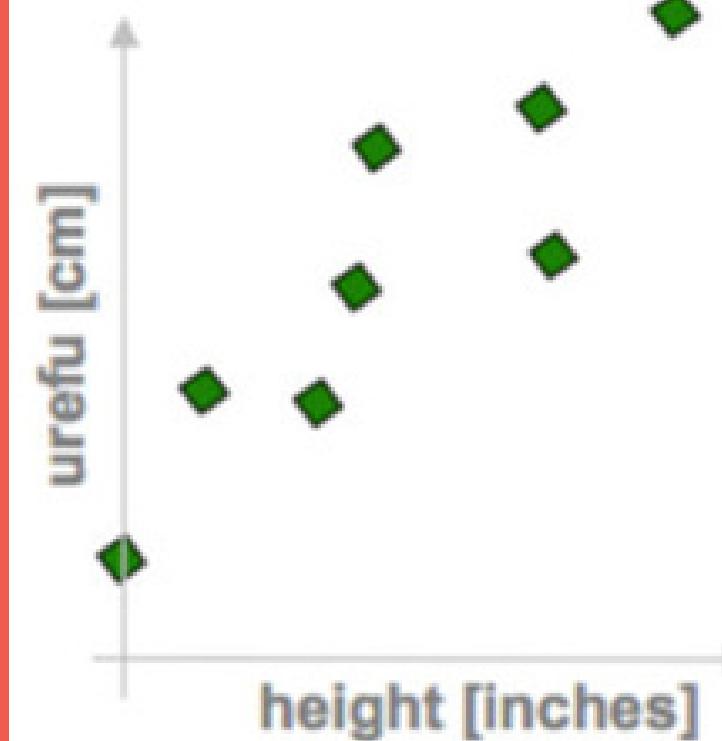


✓

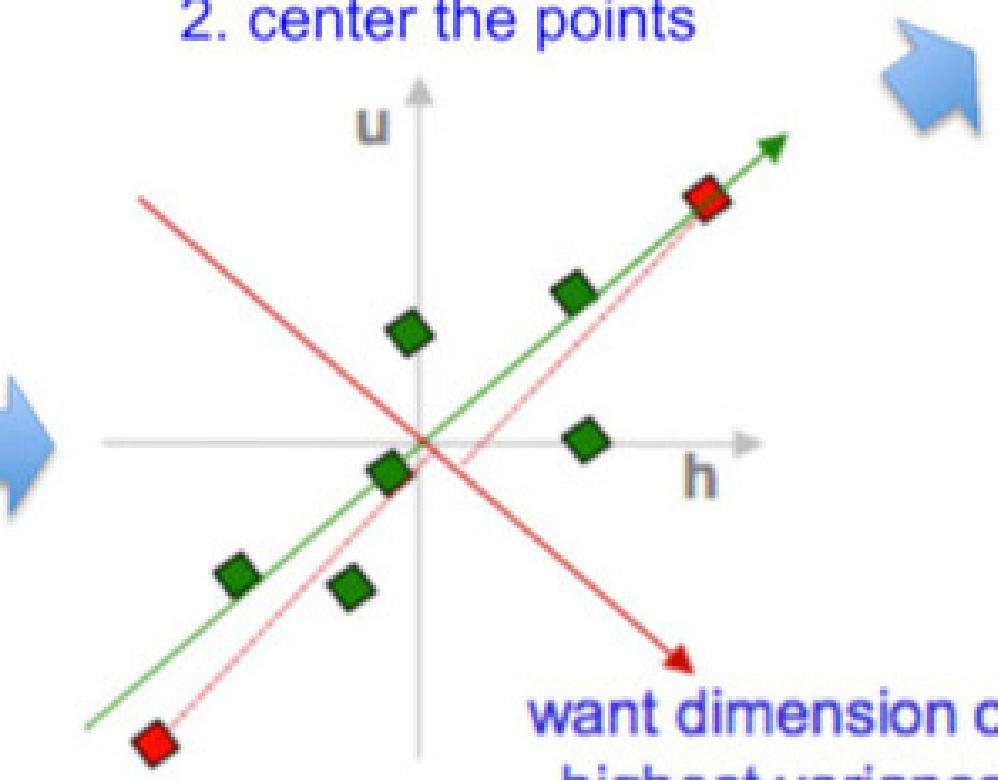
# PCA

## PCA in a nutshell

1. correlated hi-d data  
("urefu" means "height" in Swahili)



2. center the points



3. compute covariance matrix

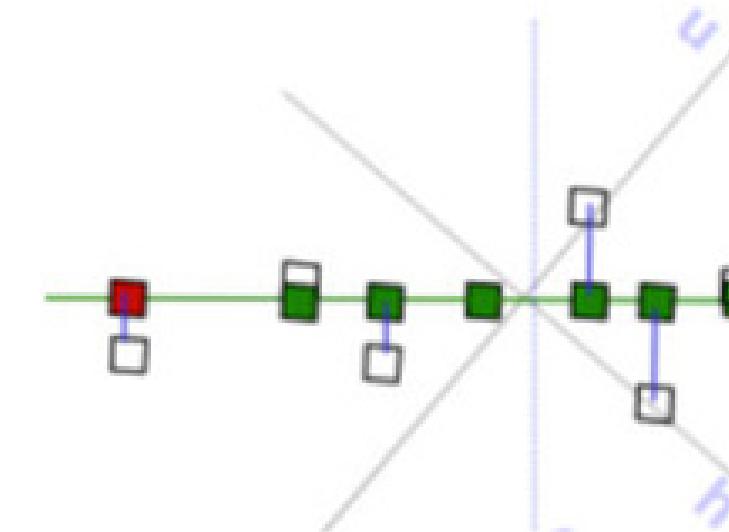
$$\begin{matrix} h & u \\ \begin{pmatrix} 2.0 & 0.8 \\ 0.8 & 0.6 \end{pmatrix} & \end{matrix} \rightarrow \text{cov}(h, u) = \frac{1}{n} \sum_{i=1}^n h_i u_i$$

4. eigenvectors + eigenvalues

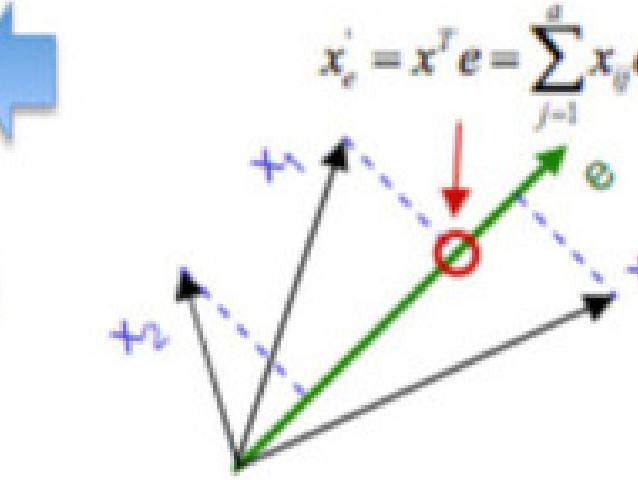
$$\begin{pmatrix} 2.0 & 0.8 \\ 0.8 & 0.6 \end{pmatrix} \begin{pmatrix} e_h \\ e_u \end{pmatrix} = \lambda_e \begin{pmatrix} e_h \\ e_u \end{pmatrix}$$
$$\begin{pmatrix} 2.0 & 0.8 \\ 0.8 & 0.6 \end{pmatrix} \begin{pmatrix} f_h \\ f_u \end{pmatrix} = \lambda_f \begin{pmatrix} f_h \\ f_u \end{pmatrix}$$

`eig(cov(data))`

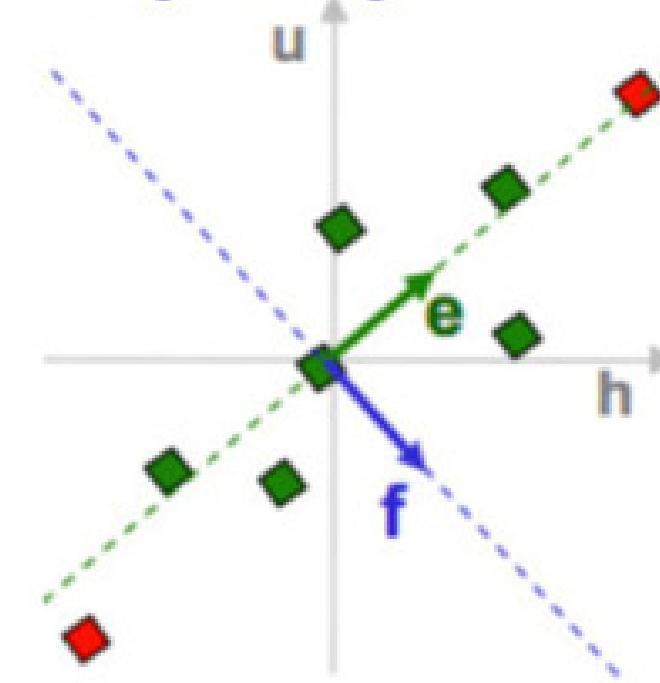
7. uncorrelated low-d data



6. project data points to those eigenvectors



5. pick m<d eigenvectors w. highest eigenvalues



# QUESTION

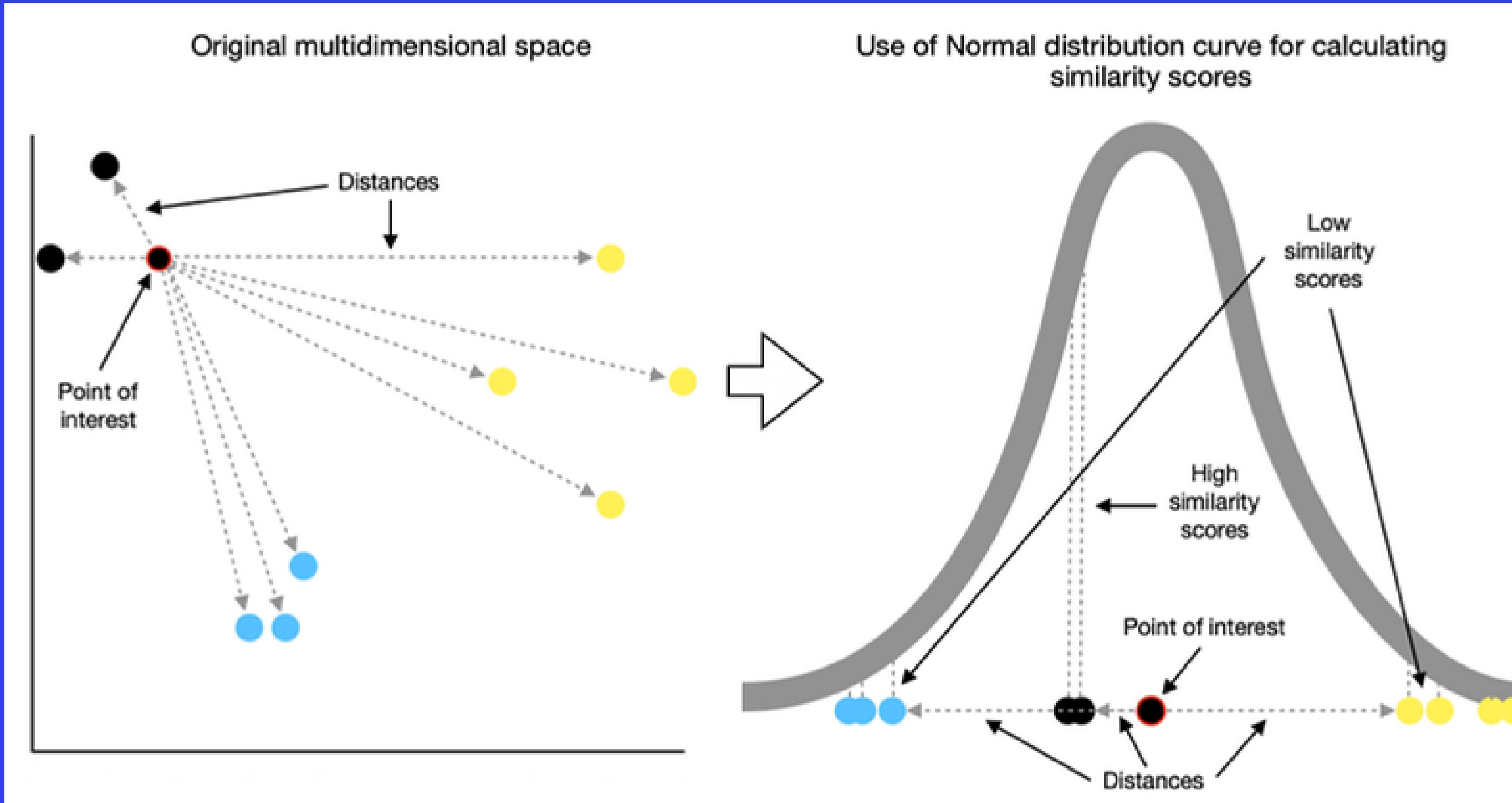
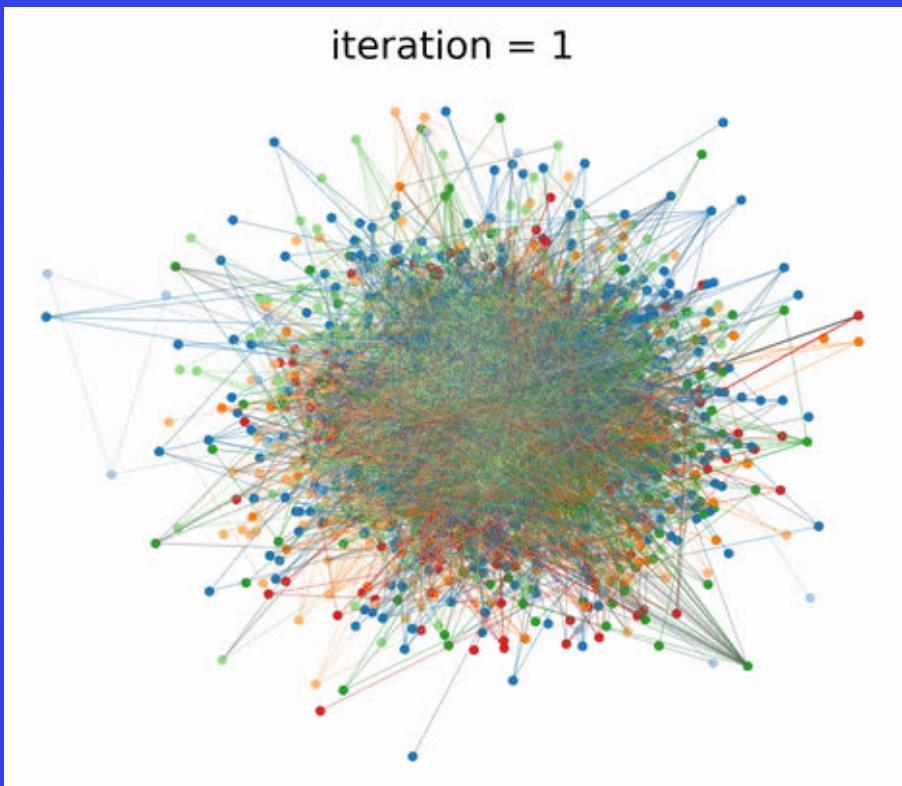
x

o

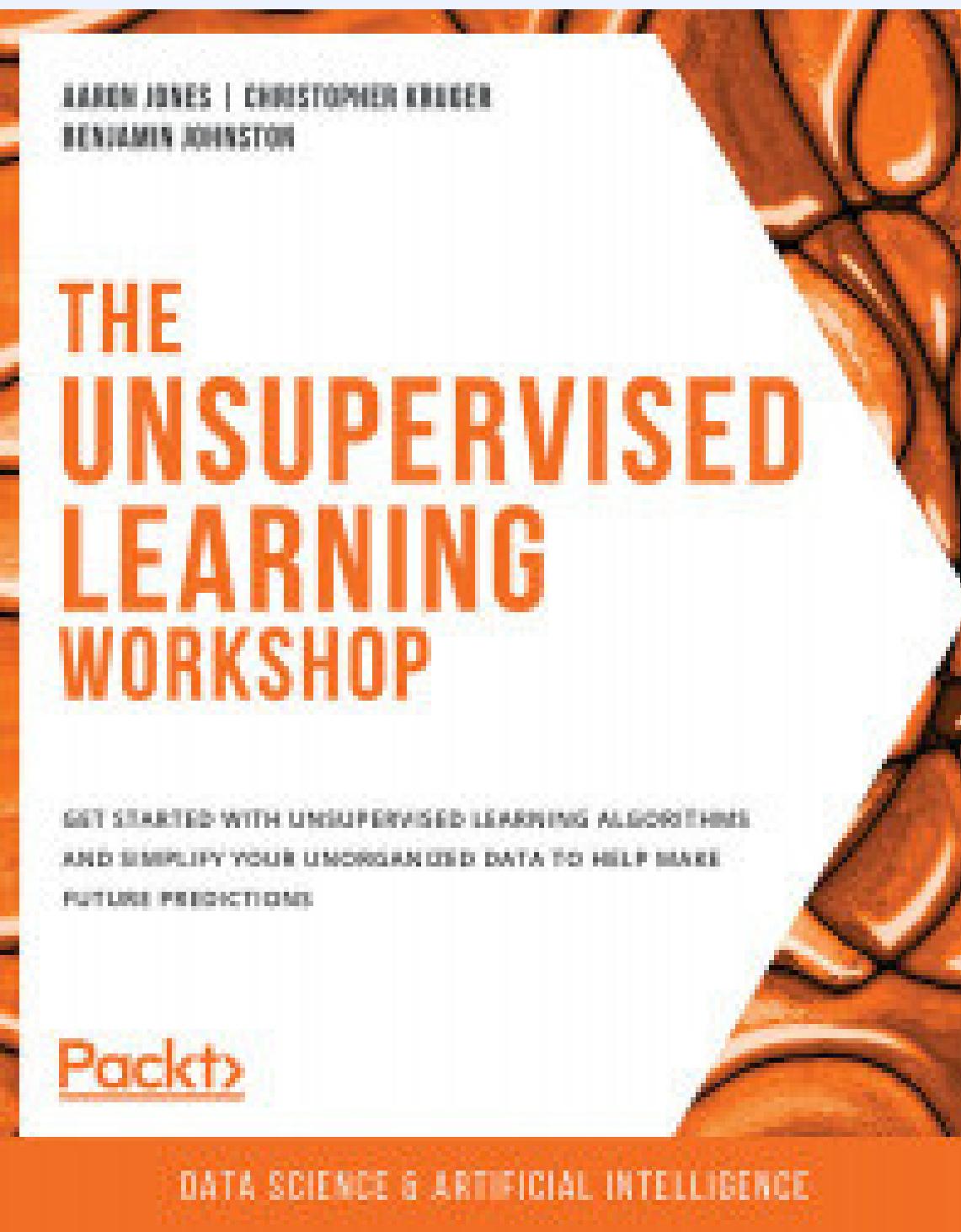
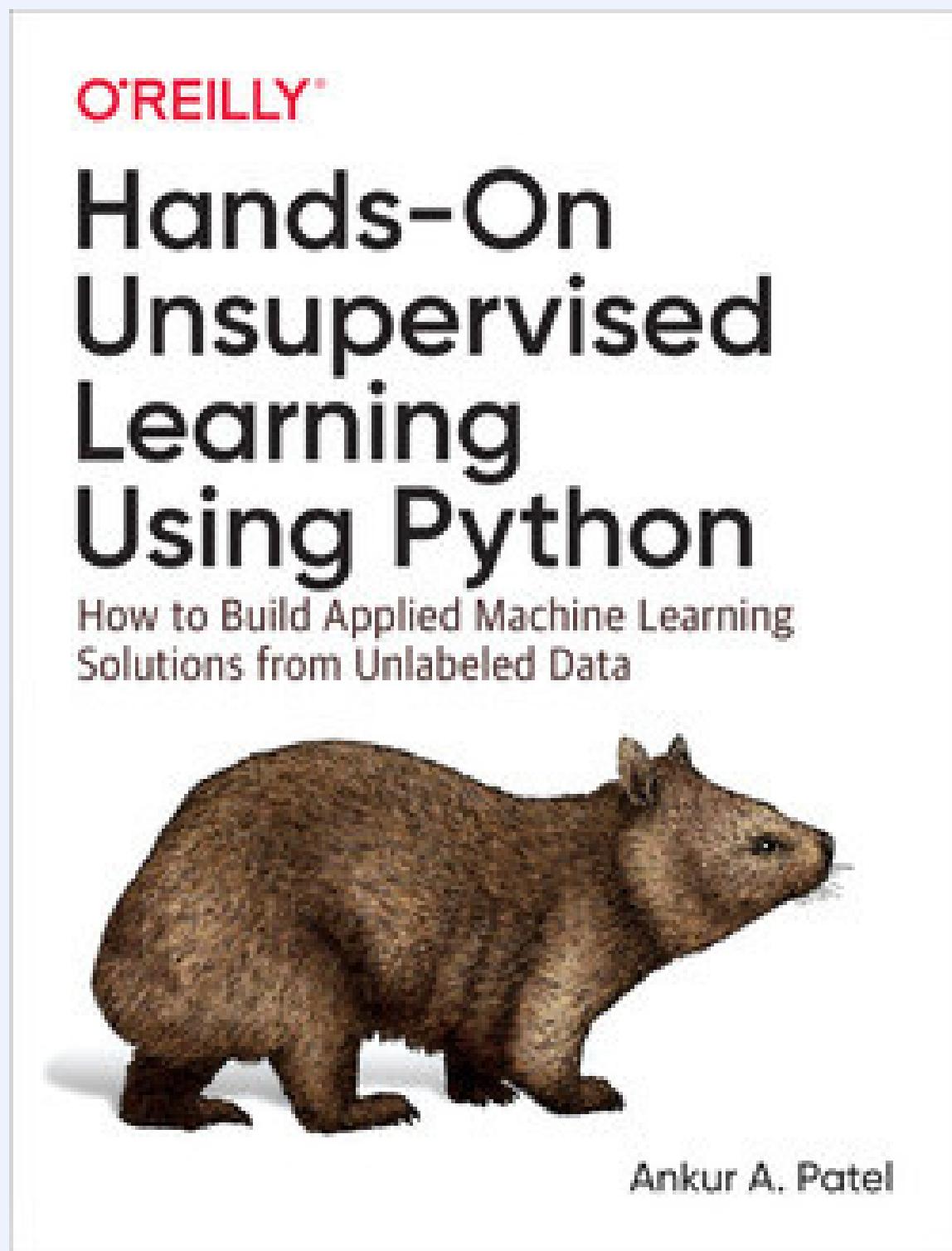
How might unsupervised learning techniques help in the exploratory data analysis stage of a data science project?



# T-SNE

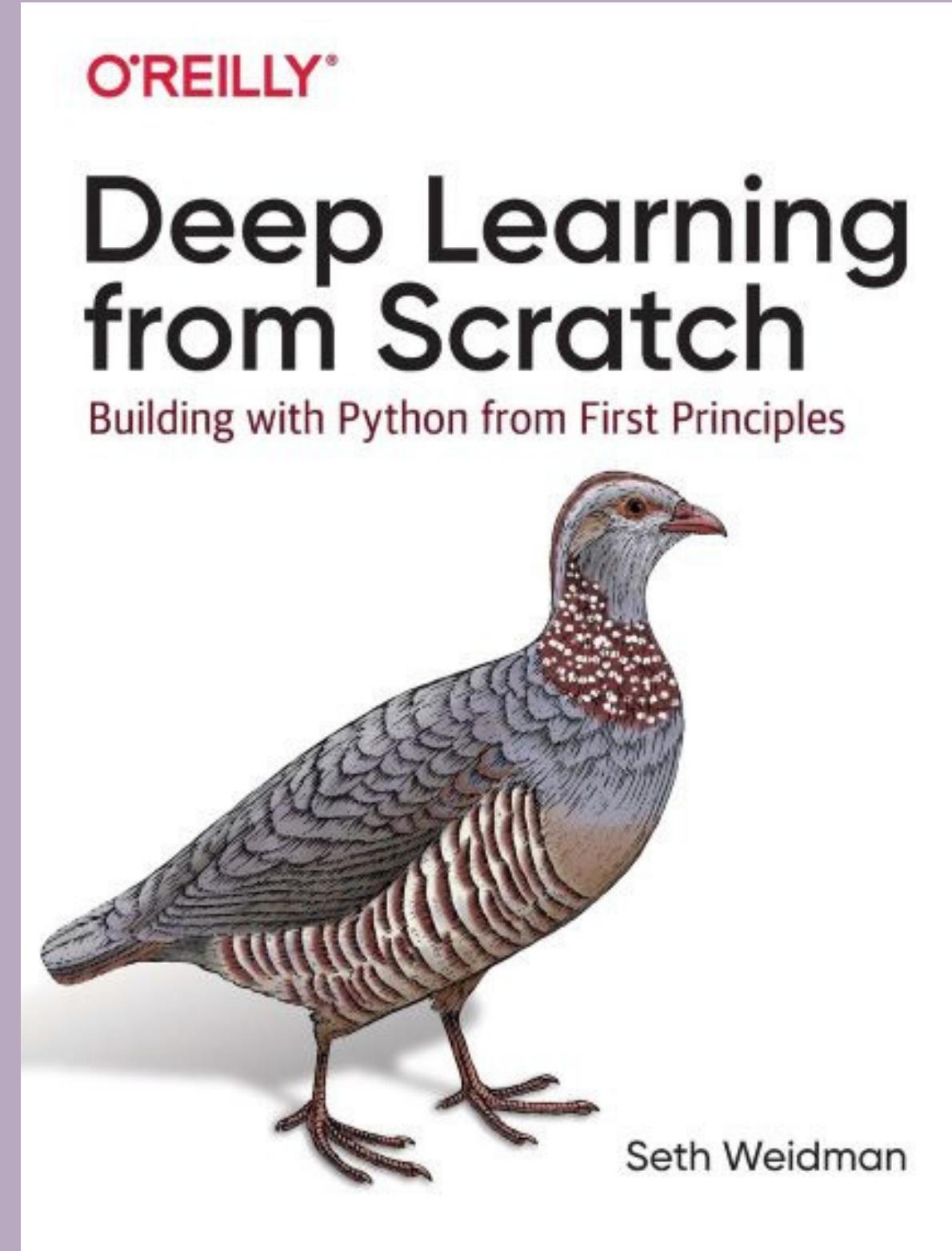
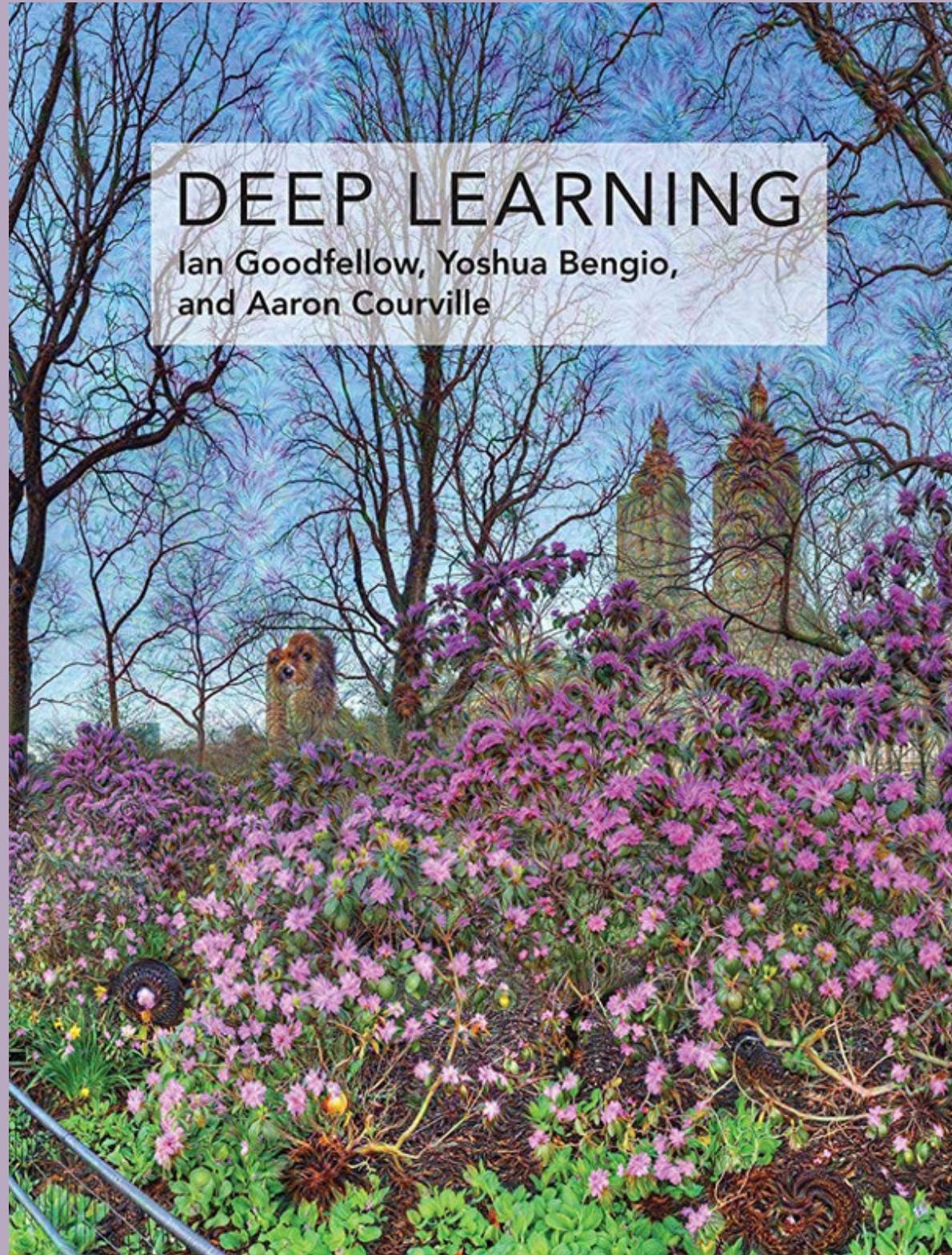


# Resources



**StatQuest with  
Joshua Starmer**

# Where to go from now !



Projects !  
Understand  
Deeply !  
Theory can be  
tiring but is  
really necessary



THANKS A LOT AND  
GOOD LUCK

