Université Constantine 2
جامعة قسنطينة 2

Module : Machine Learning (ML – SDSI)

– Course 3 –

Chapter 3 : Unsupervised Learning

Habib-Ellah GUERGOUR

Faculty of NTIC / TLSI Department

Contact: habib.guergour@univ-constantine2.dz

Université Constantine 2
جامعة قسنطينة 2

Module : Machine Learning (ML – SDSI)

– Course 3 –

Chapter 3 : Unsupervised Learning

Habib-Ellah GUERGOUR
Faculty of NTIC / TLSI Department
Contact: habib.guergour@univ-constantine2.dz

Etudiants concernés

Faculté/Institut	Département	Niveau	Spécialité
NTIC	TLSI	M1	SDSI

Goals of the Chapter

- Grasp Unsupervised Learning Basics
- Learn Key Clustering Algorithms such as **K-Means**, **Hierarchical Clustering**, and **DBSCAN**
- Assess Clustering Quality like **Silhouette Score**, **Davies-Bouldin Index**, and **Adjusted Rand Index**
- Explore Dimensionality Reduction , such as **PCA** and **t-SNE**
- Apply to Real-World Use Cases (e.g., segmentation, anomaly detection)

Main Titles

- Introduction
- Clustering
- Anomaly Detection
- Dimensionality Reduction

Introduction

Introduction

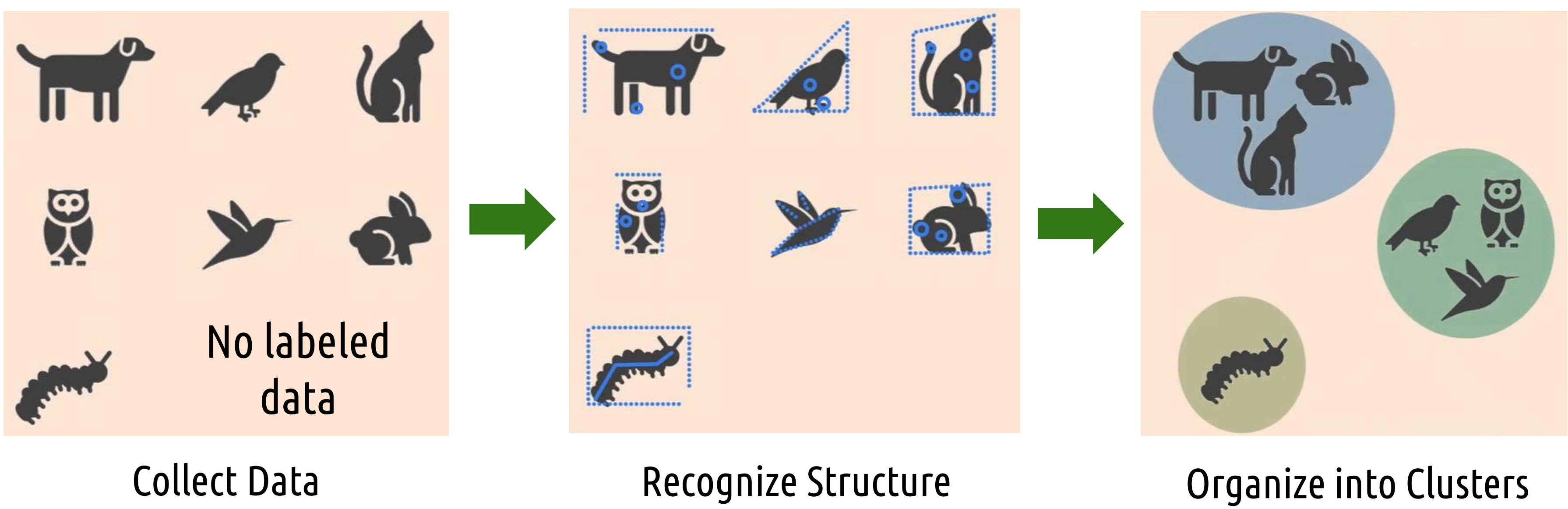
What is Unsupervised Learning?

- Unsupervised learning deals with datasets that have no labels. The goal is to discover the inherent structure in the data.
- Key Tasks:
  - **Clustering**: Group similar instances.
  - **Dimensionality Reduction**: Reduce the number of features while retaining key structure.
  - **Anomaly Detection**: Identify rare or abnormal observations.

Introduction

Clustering

- Clustering is a method to organize unlabeled data by finding natural groups based on similarities.
- Example:



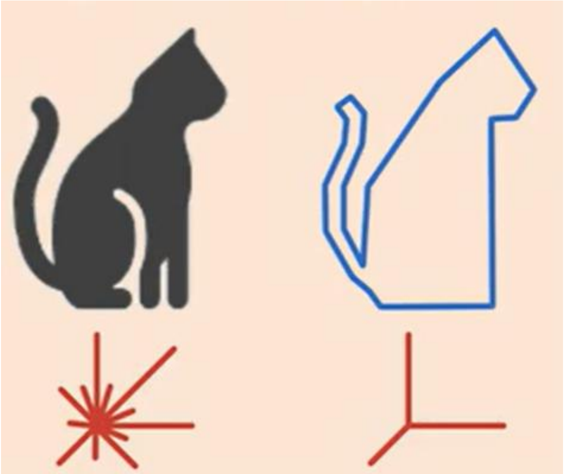
Introduction

Dimensionality Reduction

- Transform high-dimensional data into a lower-dimensional representation while retaining its meaningful structure.

Goals:

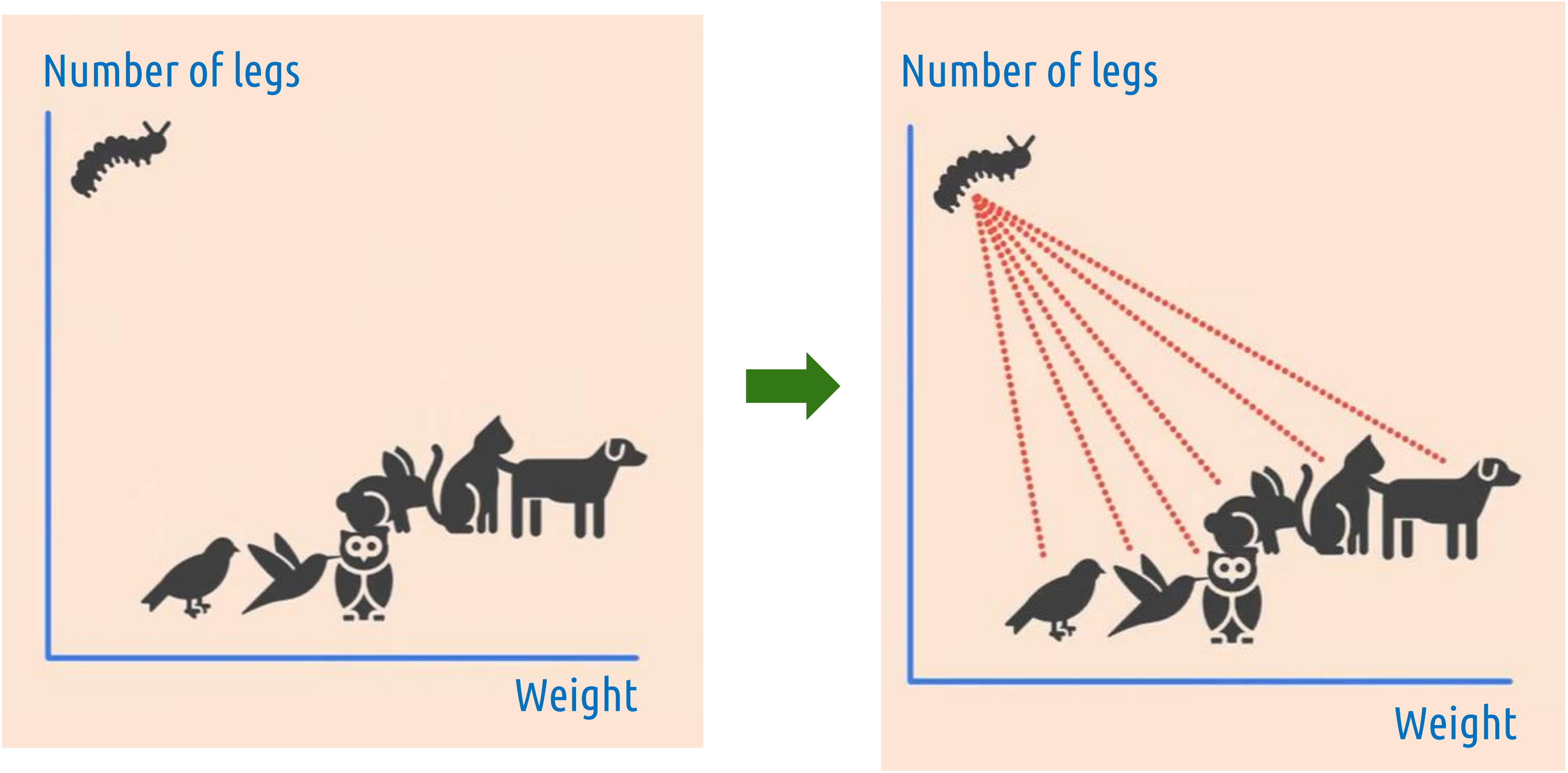
- Reducing noise (removing irrelevant features).
- Preserving critical patterns (keeping the most useful information).
- Enabling efficiency (faster computation, easier visualization).



Introduction

Anomaly Detection

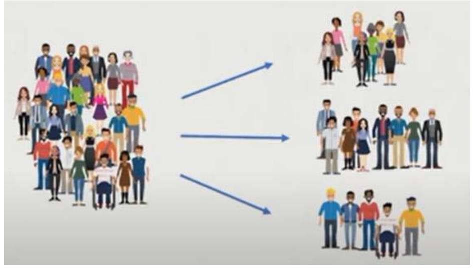
- The process of identifying rare, unusual, or suspicious data points that deviate significantly from the majority of the data



What is Unsupervised Learning?

Applications:

- Market Segmentation,
- Image Compression
- Social Network Analysis
- Fraud Detection



Clustering

What is Clustering?

- Group similar data points into clusters based on feature similarity
- No predefined labels; goal is to reveal hidden structures
- Distance/similarity metrics play a central role (e.g., Euclidean distance)
- Applications:
  - Customer segmentation
  - Document classification
  - Image grouping

Introduction

Clustering – Principal Algorithms

- **K-Means**: partitions data into K clusters based on centroid distance
- **DBSCAN**: groups data based on density; handles noise and irregular shapes
- **Hierarchical Clustering**: merges or splits clusters iteratively
- **Spectral Clustering**: uses graph Laplacian and eigen decomposition

Introduction

K-Means Clustering (Overview)

- Objective: partition n observations into K clusters
- Each observation belongs to the cluster with the nearest mean
- Assumes spherical, equally sized clusters

Introduction

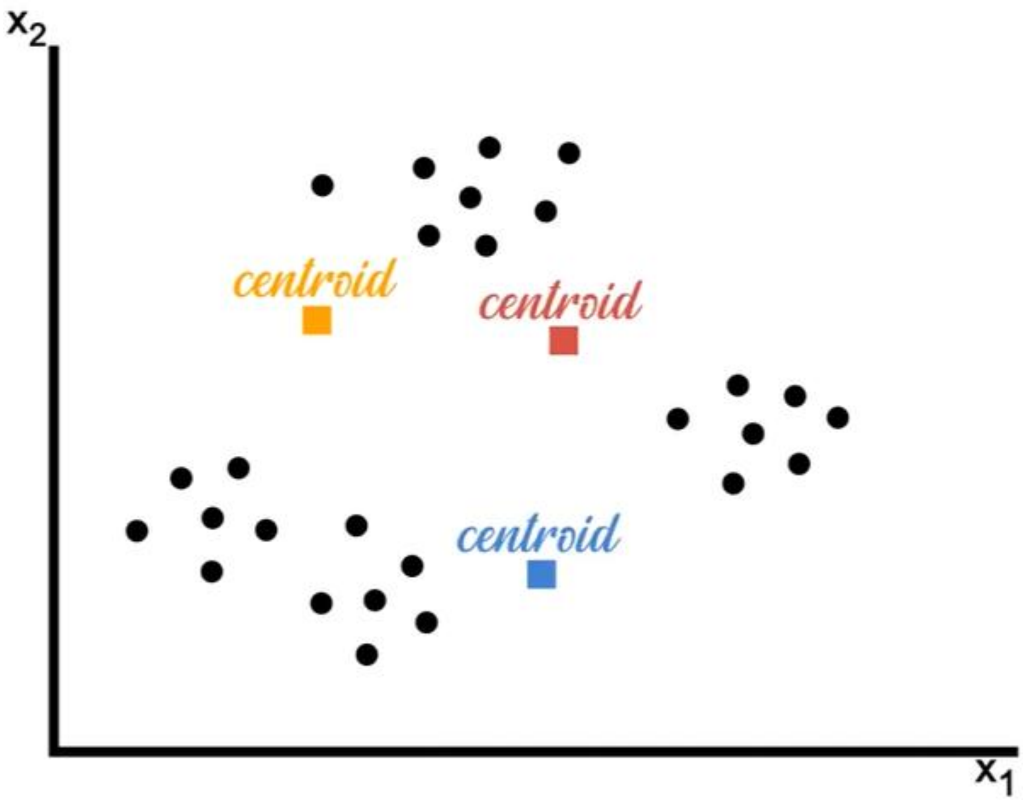
Clustering – K-means

Example

- Plot the data

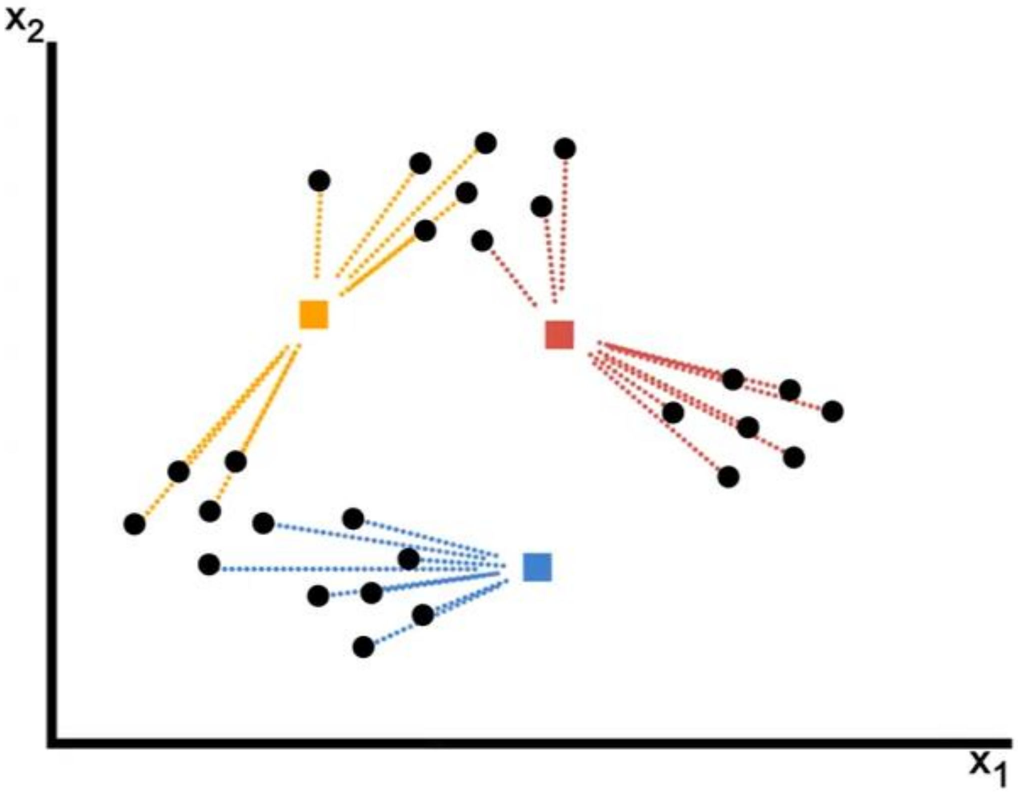
Example:

- Initialize 3 centroids randomly



Example:

- Assign each point to nearest centroid

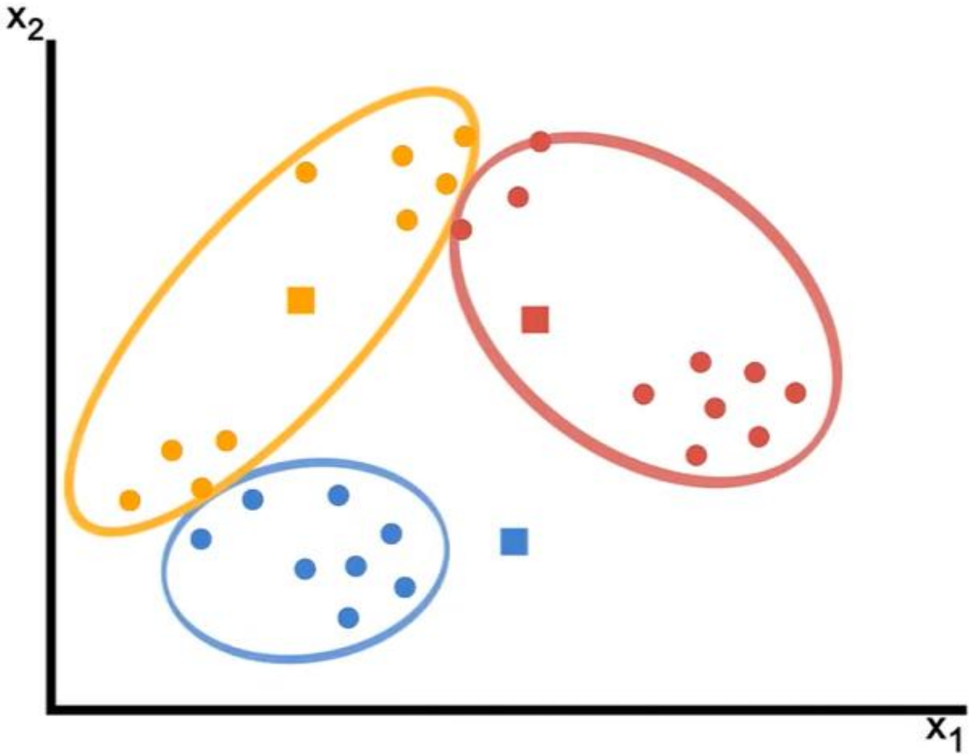


Introduction

Clustering – K-means

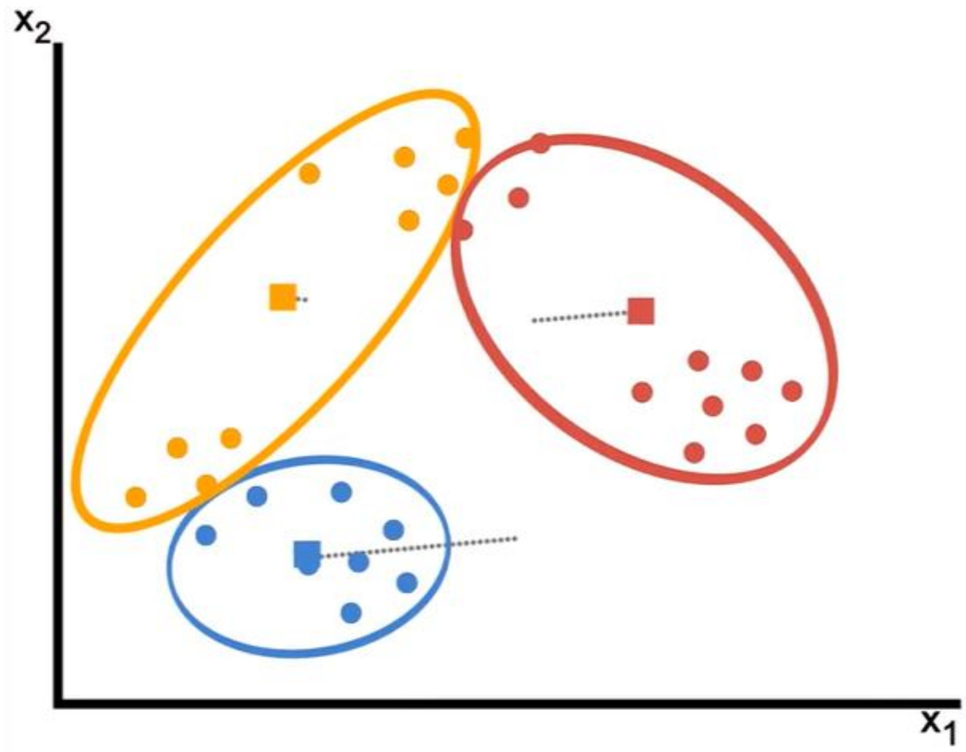
Example:

- Form the clusters



Example:

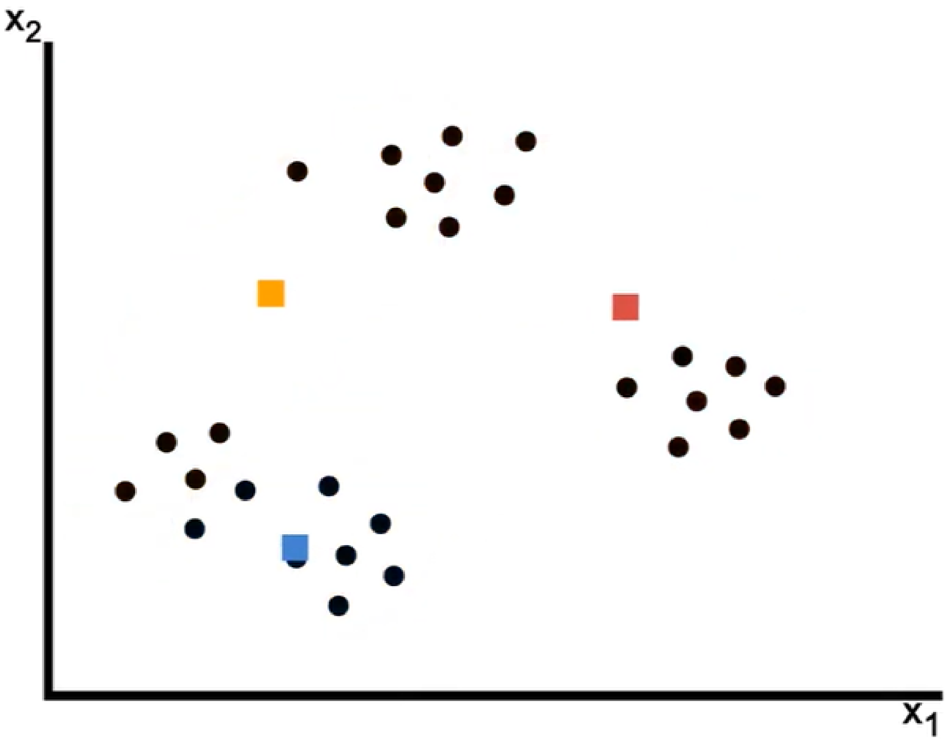
- Update centroids as mean of assigned points



Clustering – K-means

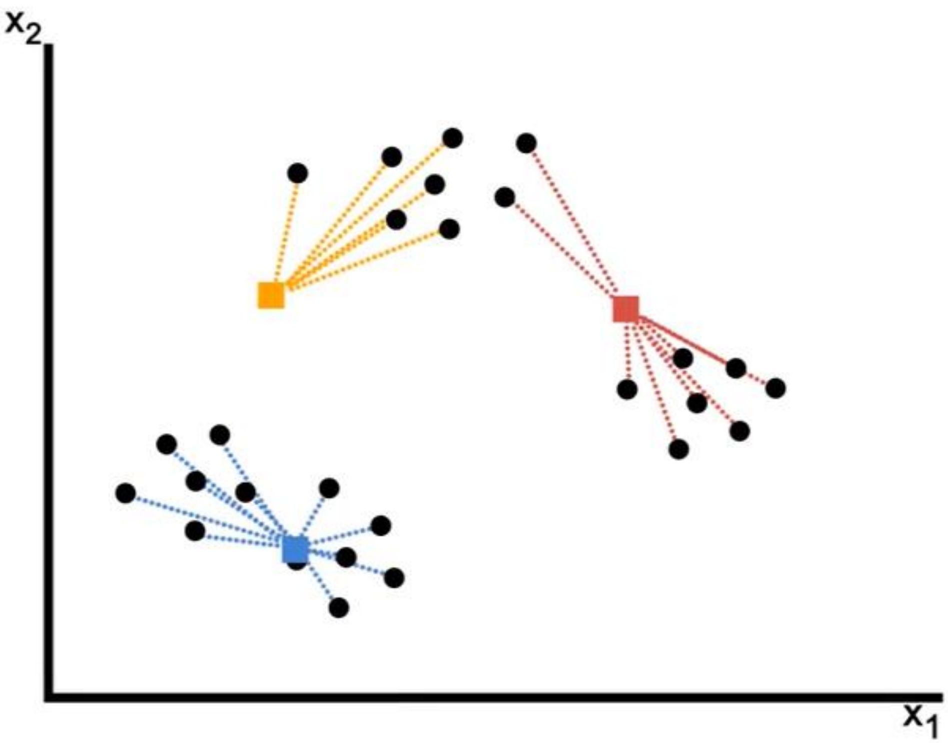
Example:

- Repeat the process until convergence



Example:

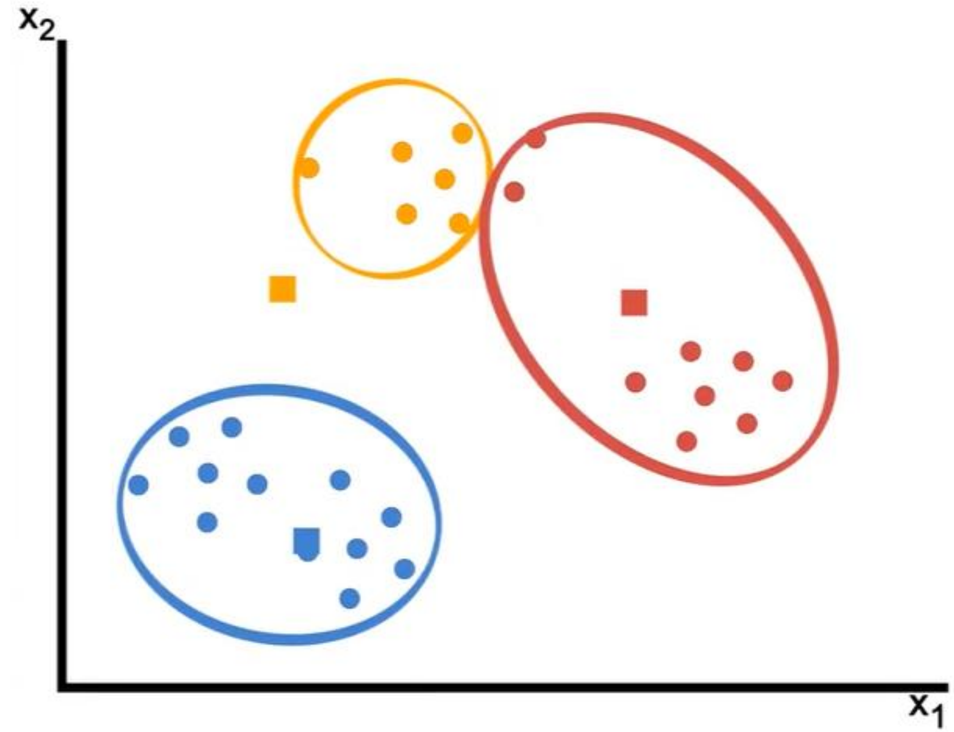
- Repeat the process until convergence



Clustering – K-means

Example:

- Repeat the process until convergence

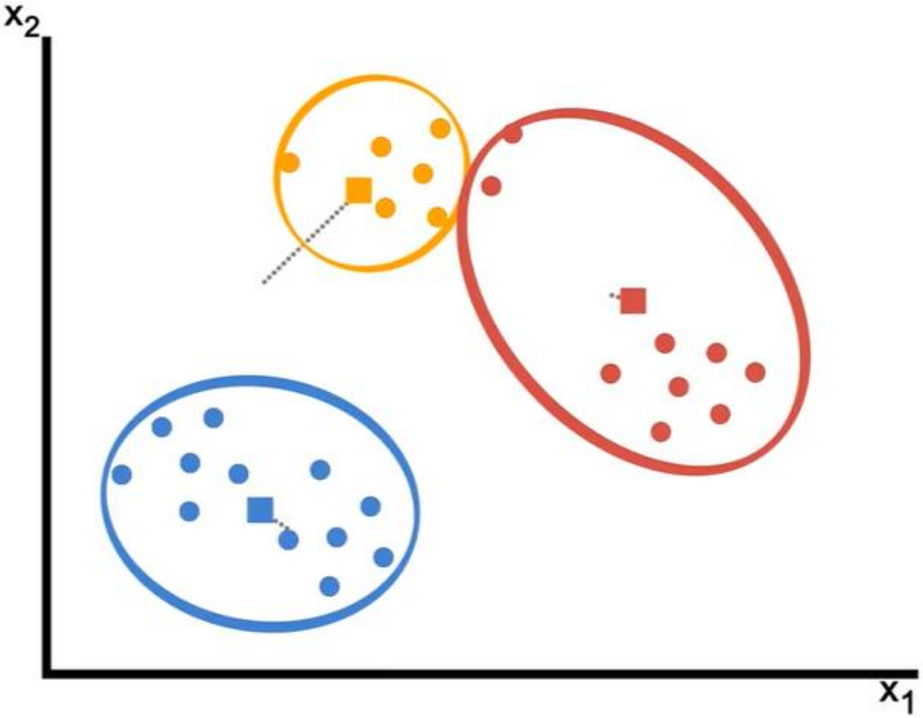


Introduction

Clustering – K-means

Example:

- Repeat the process until convergence



K-Means (Algorithm Details)

- Input: dataset X, number of clusters K
- Steps:
    1. Initialize K centroids randomly
    2. Assign each point to nearest centroid
    3. Update centroids as mean of assigned points
    4. Repeat steps 2-3 until convergence

Introduction

K-Means (Mathematical Formulation)

- Minimize within-cluster variance:

$$J = \sum_{i=1}^{n} \sum_{k=1}^{K} r_{ik} \|x_i - \mu_k\|^2$$

Where:

- r_{ik} = 1 if point x_i belongs to cluster k, 0 otherwise
- μ_k: centroid of cluster k

Introduction

Evaluating Clustering (Inertia & Silhouette)

Inertia:

- sum of squared distances of samples to their closest cluster center
- Measures compactness of clusters
- Lower inertia indicates tighter clusters

Silhouette Score:

- Measures cohesion vs. separation
- Values range from -1 to 1
- High value means points are well matched within their cluster and poorly matched to others

Choosing the Number of Clusters (K)

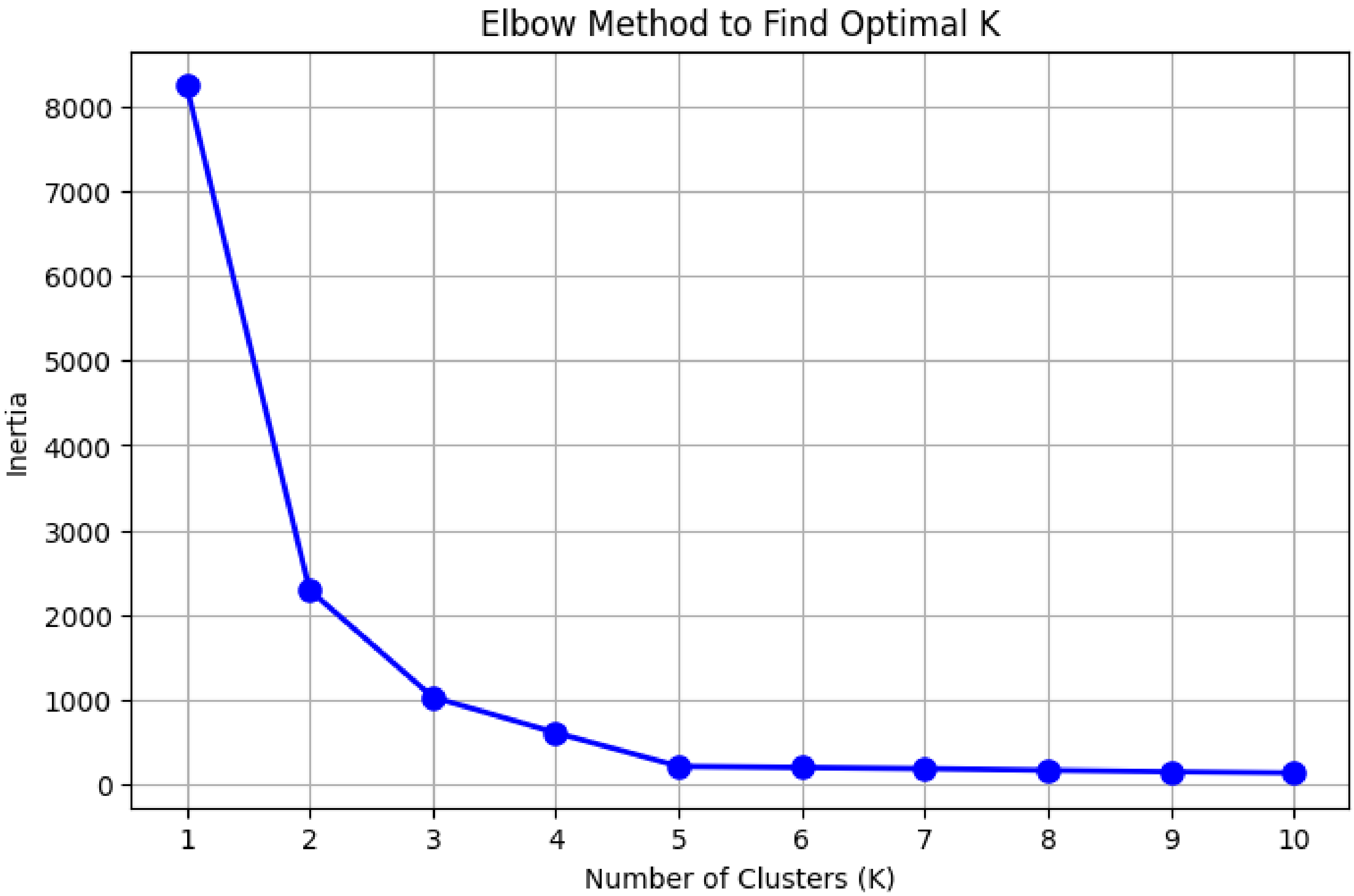
- Choosing an appropriate K is crucial for optimal clustering.
- Common methods:
  - Elbow Method
  - Silhouette Score
  - Gap Statistic

Choosing the Number of Clusters (K)

Elbow Method:

- Plot inertia vs. number of clusters.
- Look for a "knee" point where inertia starts decreasing more slowly.
- Suggests optimal trade-off between number of clusters and compactness.

Elbow Method (Example)



Anomaly Detection

Anomaly Detection

What is it?

- Anomaly detection is the process of identifying rare, unusual, or suspicious data points that deviate significantly from the majority of a dataset. These anomalies (also called outliers) can indicate:
  - Fraud (e.g., credit card transactions)
  - Defects (e.g., faulty sensors in manufacturing)
  - Cyberattacks (e.g., unusual network traffic)
  - Biological abnormalities (e.g., rare diseases in medical data)

Anomaly Detection

Anomaly Detection Techniques

Several approaches exist, each suited to different scenarios:

- Isolation Forest → Fast, tree-based outlier isolation
- Statistical Methods (Z-Score) → Simple threshold-based detection
- Density-Based (LOF, DBSCAN) → Neighborhood density analysis
- Boundary-Based (One-Class SVM) → Learned normal-data boundaries
- ...

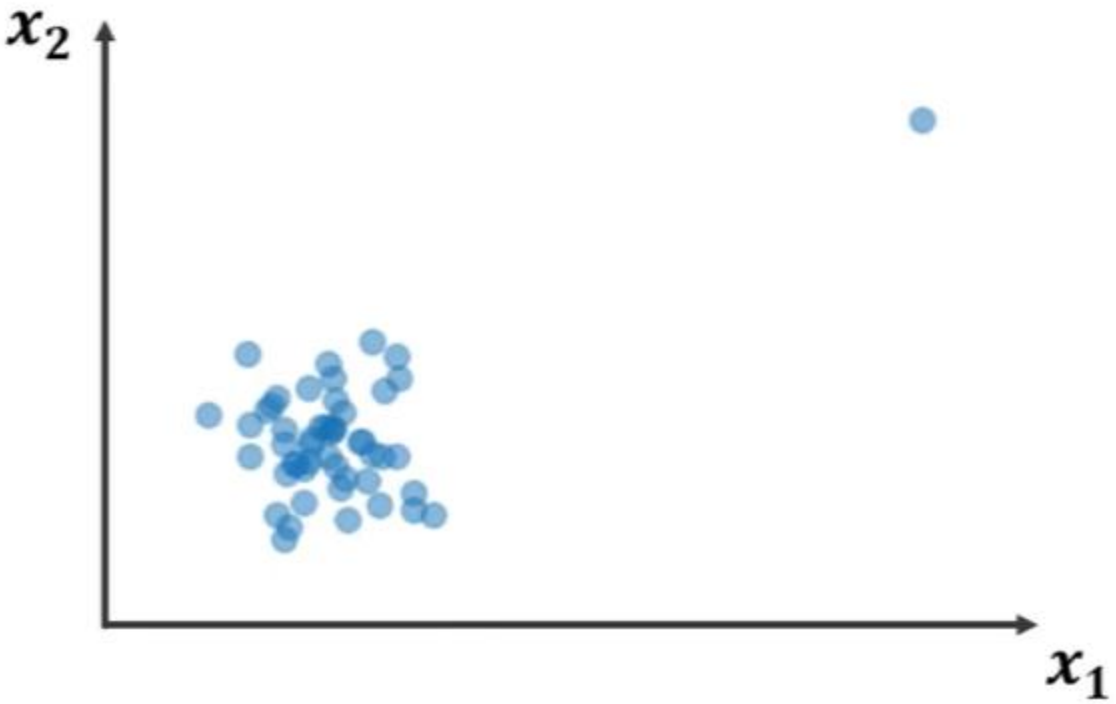
s) → Neural network

Isolation Forest

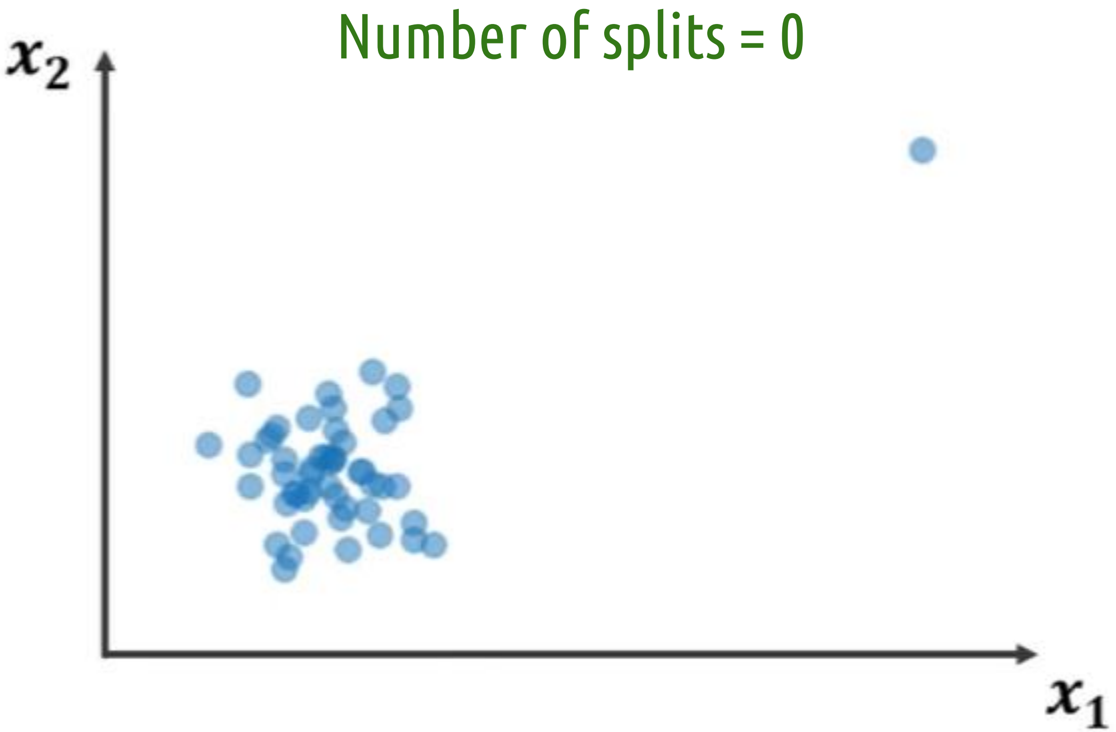
How it works:

- Builds random decision trees to "isolate" points
- Anomalies require fewer splits to separate
- Best when: You need speed and have high-dimensional data
- Limitation: May miss grouped anomalies

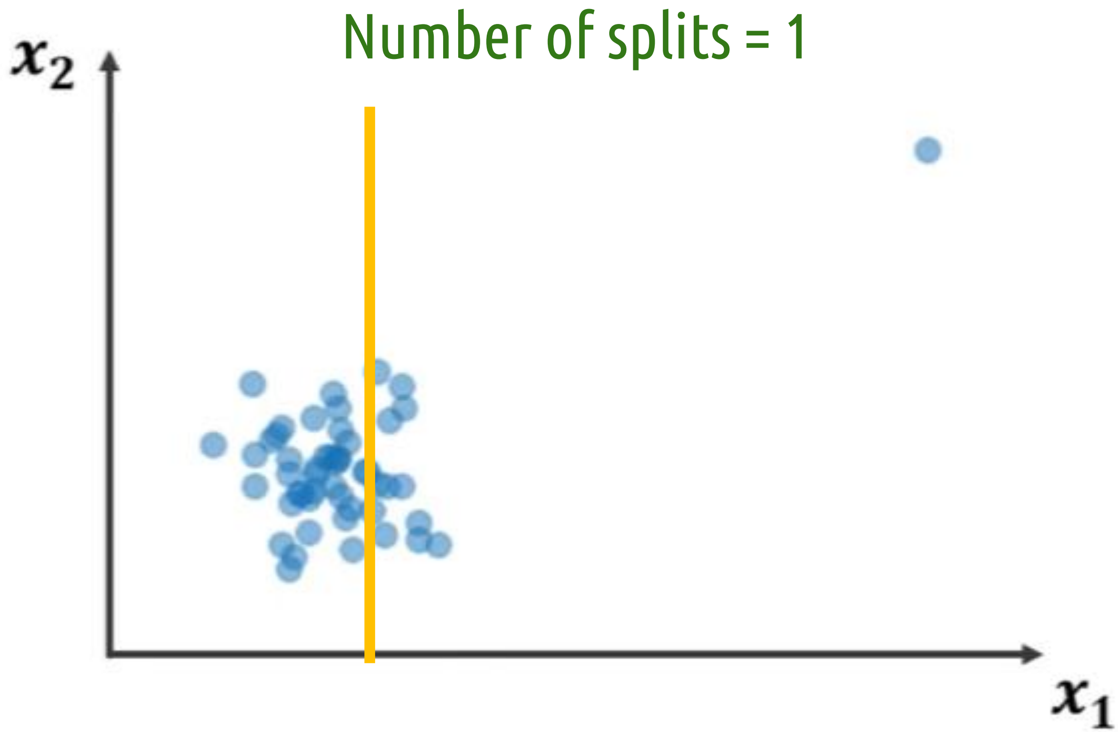
Example



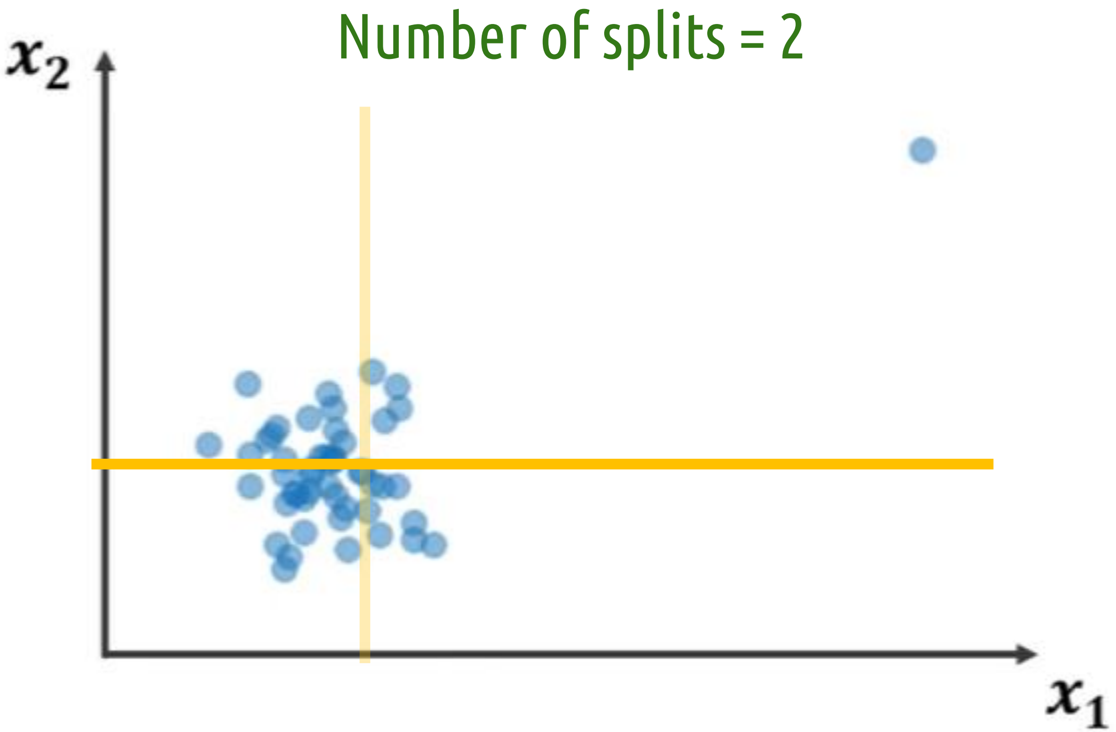
Example



Example

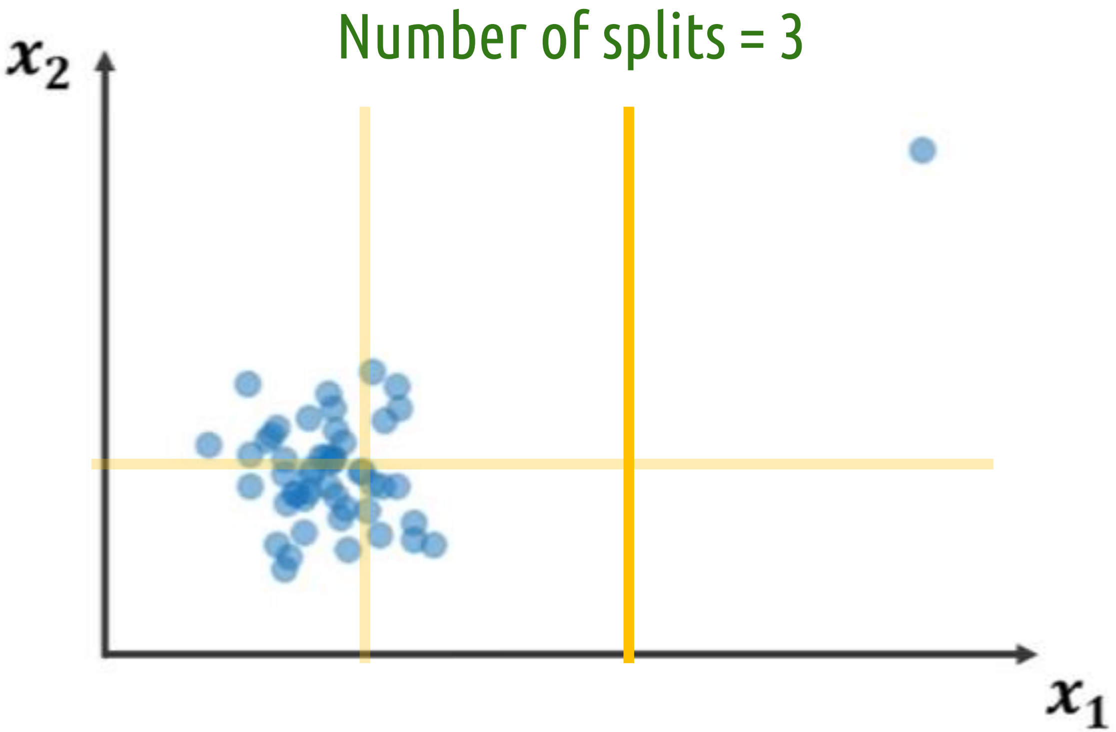


Example

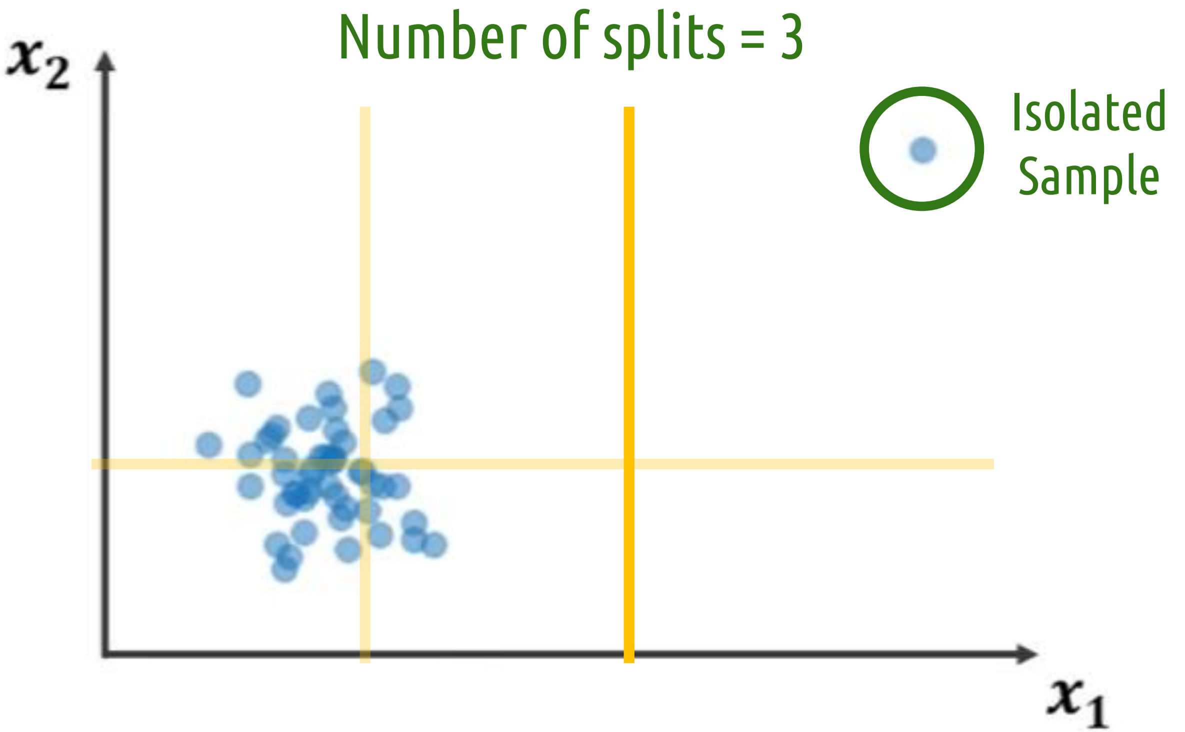


Isolation Forest

Example



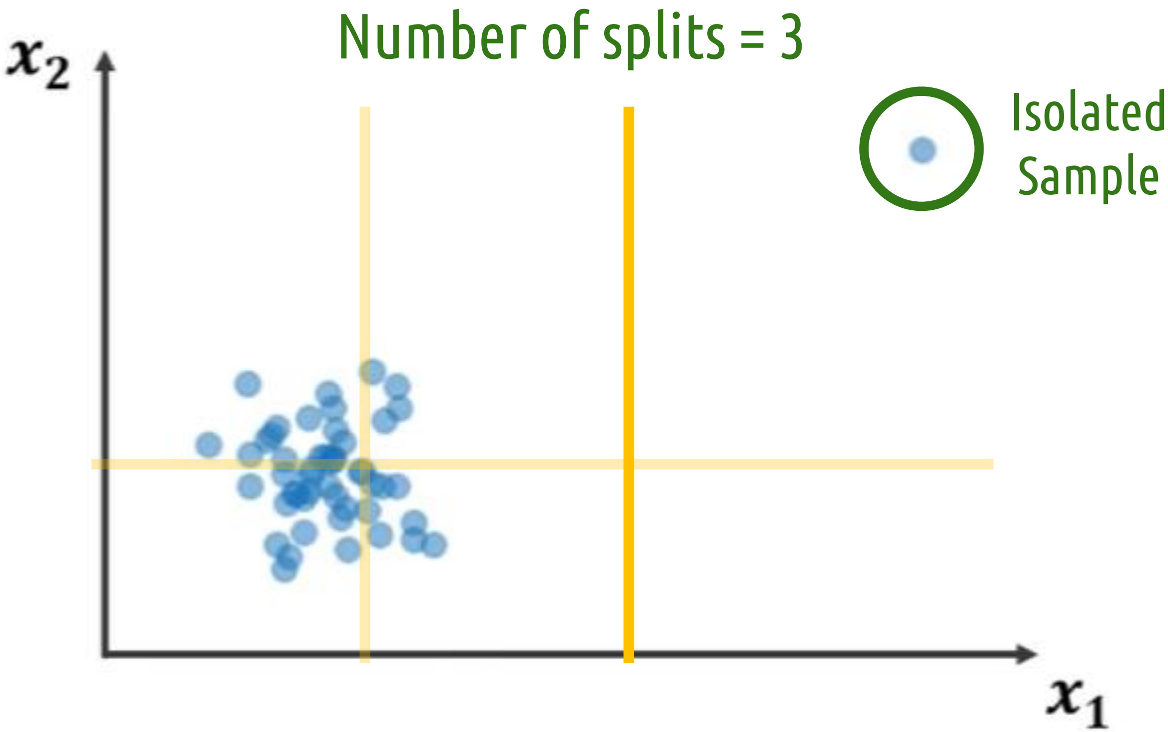
Example



Anomaly Detection

Isolation Forest

Example



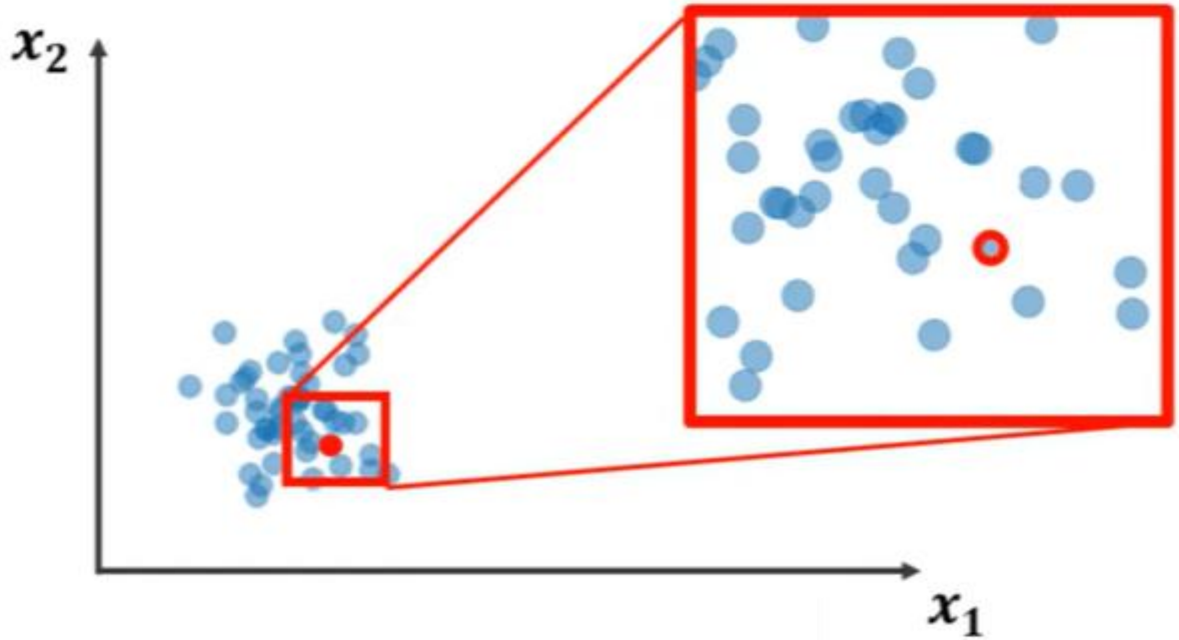
Low number of splits = High probability of anomaly

Anomaly Detection

Isolation Forest

Example

By comparison, isolating a sample lost within the mass...



Requires very high number of splits

Dimensionality Reduction

Dimensionality Reduction

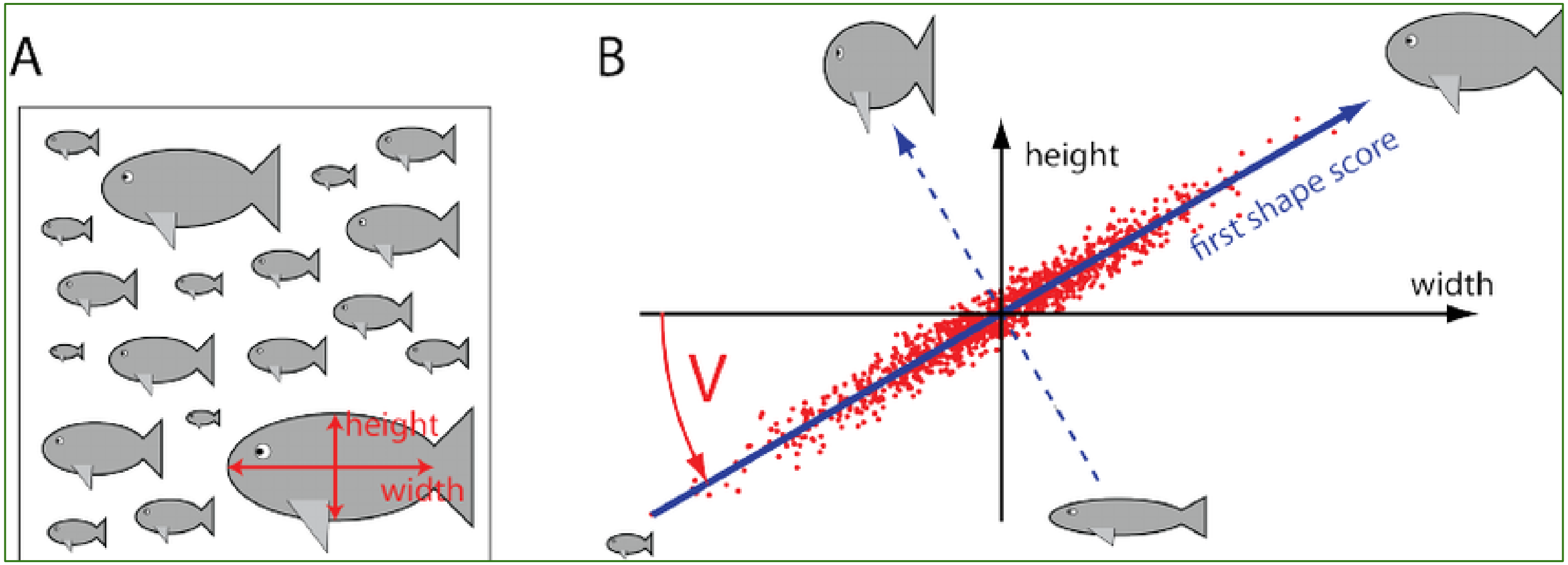
Principles

- Reduce number of features while preserving essential structure
- Removes redundancy and noise, improves visualization and learning
- **Main methods:**
    - PCA (Principal Component Analysis)
    - t-SNE, UMAP, Autoencoders (used for non-linear manifolds)

Dimensionality Reduction

PCA – Principal Component Analysis (Overview)

- Linear dimensionality reduction method
- Projects data onto orthogonal directions (principal components)
- Directions maximize variance of the data



Principal Steps of PCA:

1) Standardize the data (Optional but recommended: center and scale the features so they have mean 0 and variance 1)

2) Compute the covariance matrix

$$\boxed{\mathbf{S} = \frac{1}{n-1} X^T X}$$

3) Compute eigenvalues and eigenvectors of the covariance matrix. Solve:

$$\boxed{\mathbf{S}\mathbf{w} = \lambda \mathbf{w}}$$

Principal Steps of PCA:

4) Sort eigenvalues and eigenvectors
   - Sort the eigenvalues λ from largest to smallest.
   - Keep the corresponding eigenvectors w in the same order.

5) Select the top k eigenvectors (based on how much variance you want to keep).

6) Project the data onto the new space

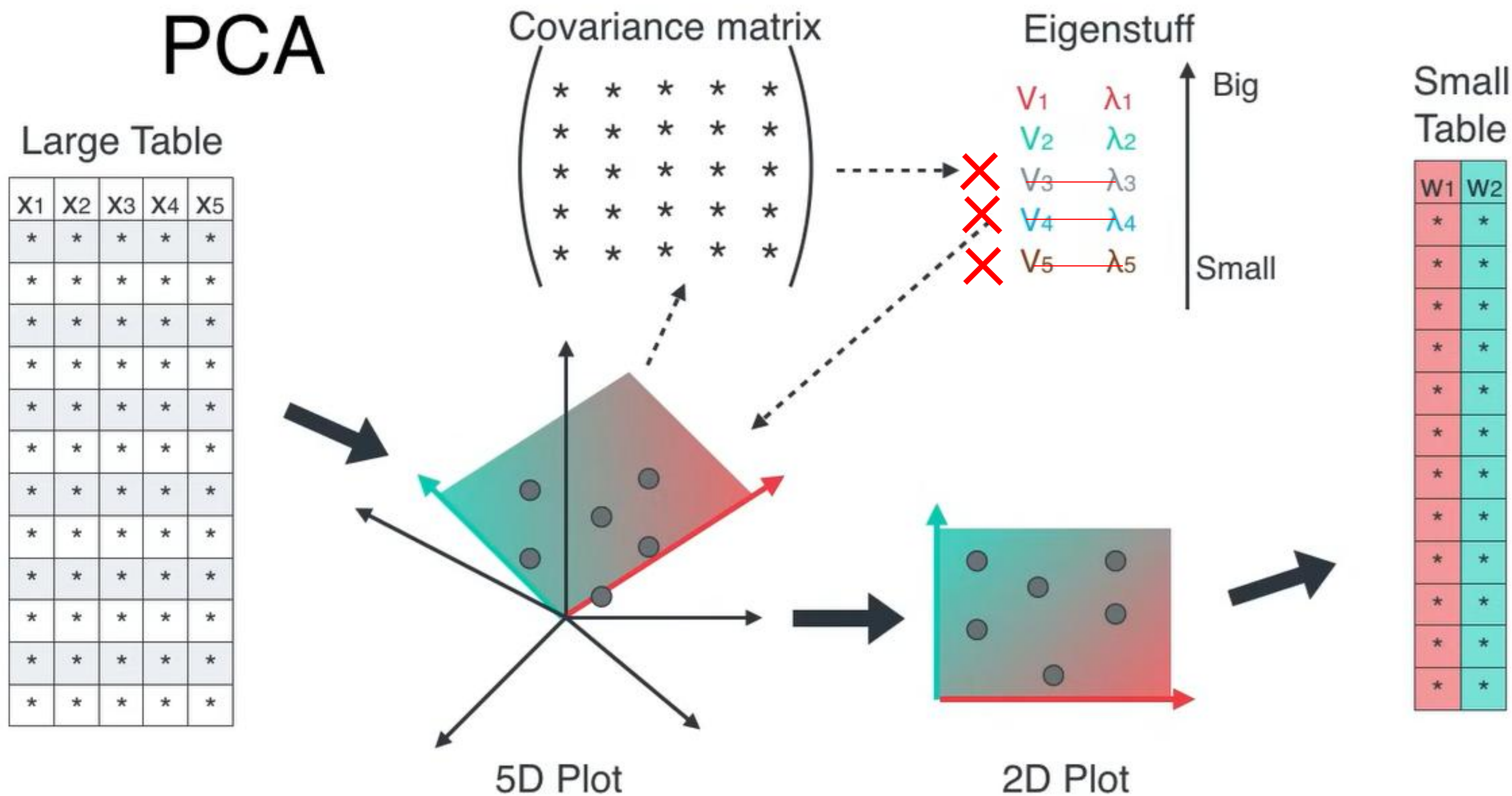
   New data:

$$\boxed{X_{\text{new}} = X \times W_k}$$

   where W_k is the matrix of the top k eigenvectors.

Dimensionality Reduction

PCA – Principal Component Analysis (Example)



Dimensionality Reduction

How to choose the number of components?

Variance explained threshold (most common way):

- Choose k so that the cumulative variance is above a threshold, like 95% or 99%.
- Formula:

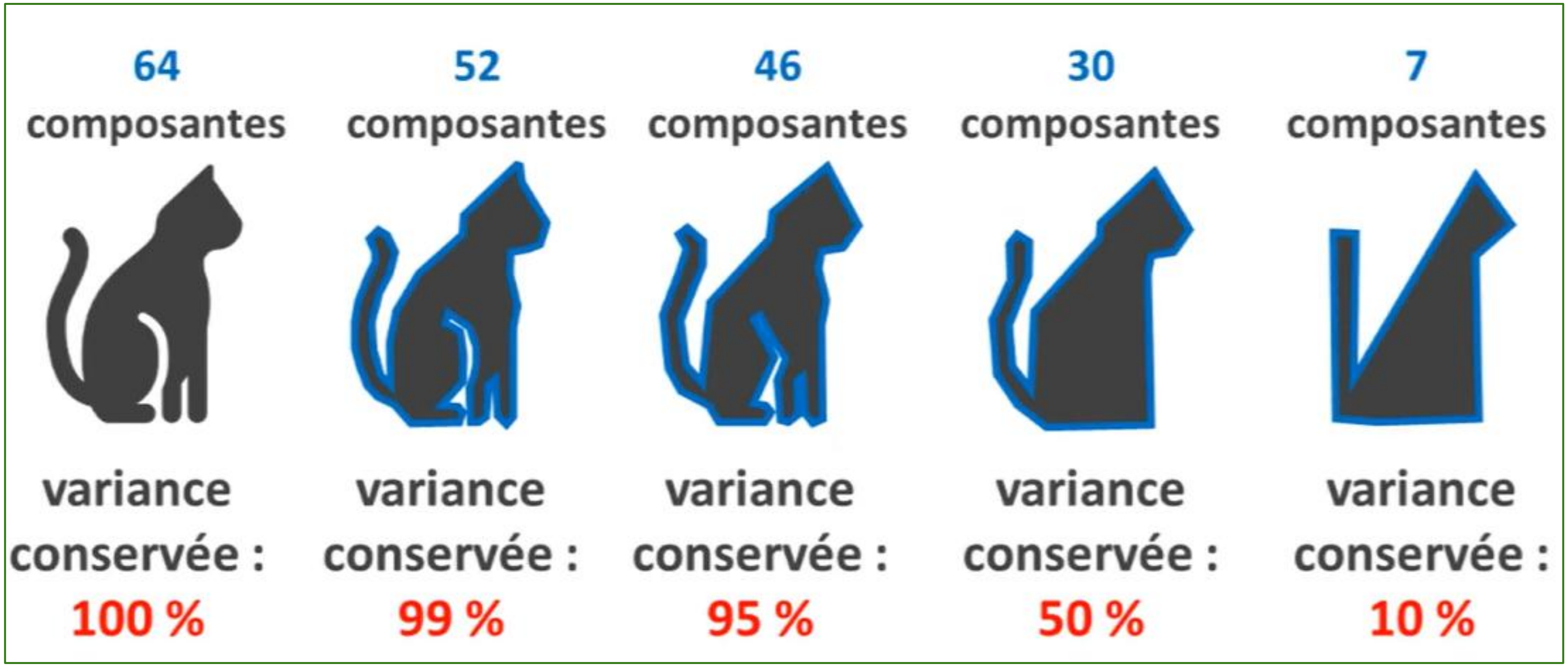
$$\boxed{\text{Cumulative variance} = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{n} \lambda_i}}$$

Here, λi are eigenvalues sorted decreasingly.

Dimensionality Reduction

How to choose the number of components?

- Example:



How to choose the number of components?

Other methods:

- Scree plot (elbow method)
- Parallel Analysis
- Cross-validation
- …

Reference

Reference

- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* Springer. A comprehensive textbook with in-depth coverage of clustering and dimensionality reduction.
- Ng, A. (2025). *Unsupervised Learning, Recommenders, Reinforcement Learning* [https://www.coursera.org/learn/unsupervised-learning-recommenders-reinforcement-learning]. Coursera.
- Saint-Cirgue, G. [Machine Learnia]. (2025). *Apprentissage non-supervisé avec Python* (24/30) [Video]. YouTube. https://www.youtube.com/watch?v=FTtzd31IAOw&t=1546s

Thank you for your attention...