

Deep Unsupervised Learning (Overview)

Naeemullah Khan

naeemullah.khan@kaust.edu.sa



جامعة الملك عبد الله
للعلوم والتقنية
King Abdullah University of
Science and Technology

KAUST Academy
King Abdullah University of Science and Technology

May 21, 2025

1. Definition
2. Applications
3. Challenges
4. Types
 - 4.1 Clustering
 - 4.2 Dimensionality Reduction
 - 4.3 Anomaly Detection
5. Dimensionality Reduction
 - 5.1 PCA
 - 5.2 t-SNE
 - 5.3 UMAP

- ▶ We have a dataset without labels. Our goal is to learn something interesting about the underlying structure of the data:
 - Clusters hidden in the dataset.
 - Outliers: particularly unusual and/or interesting data points.
 - Useful signals hidden in the noise, e.g., human speech over a noisy background.

- ▶ **Data:** Unlabeled data, e.g., images, text, or sensor readings.
- ▶ **Model:** A mathematical representation of the data, e.g., a mixture model or a neural network.
- ▶ **Objective function:** A measure of how well the model fits the data, e.g., likelihood or reconstruction error.
- ▶ **Optimization algorithm:** An algorithm to minimize the objective function, e.g., gradient descent or expectation-maximization.
- ▶ **Evaluation metrics:** Measures to assess the quality of the learned model, e.g., silhouette score or clustering accuracy.
- ▶ **Applications:** Use cases for unsupervised learning, e.g., clustering, dimensionality reduction, or anomaly detection.

Aspect	Supervised Learning	Unsupervised Learning
Objective	Learn a function f from labeled input-output pairs.	Discover structure or representations in unlabeled data.
Evaluation	Accuracy, precision/recall on held-out labels.	Clustering validity indices (e.g. silhouette), reconstruction error.
Cost	Methods range from $\mathcal{O}(n)$ to $\mathcal{O}(n^3)$ per fit.	k-means $\mathcal{O}(nkd)$, hierarchical $\mathcal{O}(n^2)$, PCA $\mathcal{O}(nd^2)$.
Labels/Clusters	Fixed, known set of classes.	Number of clusters unknown; must be chosen or inferred.
Output	Classifier or regressor for new inputs.	Cluster assignments, embeddings, density models, or generative samples.

Table 1: Key differences between Supervised and Unsupervised Learning

Unsupervised learning is used in various fields and applications, including:

- ▶ **Visualisation:** Identifying and making accessible useful hidden structures in the data.
- ▶ **Anomaly Detection:** Identifying faulty components that are likely to break soon.
- ▶ **Signal denoising:** Extracting human speech from a noisy recording.
- ▶ **Generative Models:** Learning to generate new data points similar to the training data.
- ▶ **Feature Learning:** Automatically discovering useful representations of the data.
- ▶ **Data Preprocessing:** Cleaning and transforming data for better performance in supervised learning tasks.

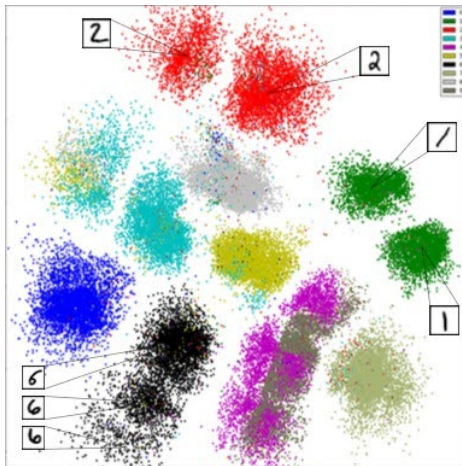


Figure 2: Unsupervised learning can discover structure in digits without any labels.

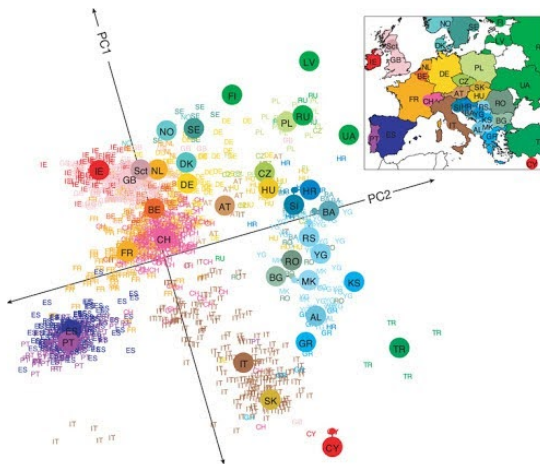


Figure 3: Dimensionality reduction applied to DNA reveal the geography of European countries.

What is Deep Unsupervised Learning?

What is Deep Unsupervised Learning? (cont.)

- ▶ Capturing rich patterns in raw data with deep networks in a label-free way.

- ▶ Capturing rich patterns in raw data with deep networks in a label-free way.
 - **Generative Models:** Recreate raw data distribution.