Deep Unsupervised Learning (Overview)

Naeemullah Khan
naeemullah.khan@kaust.edu.sa

KAUST Academy
King Abdullah University of Science and Technology

May 20, 2025

Table of Contents

Unsupervised Learning - Definition

- We have a dataset wihtout labels. Out 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.

Components of Unsupervised Learning

- **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.

Supervised vs Unsupervised Learning

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