

Homework 4

Antoine Legendre

November 2024

1 Exercise 1

Machine learning is a branch of artificial intelligence that focuses on building systems that can learn from data and improve their performance over time without being explicitly programmed. It allows computers to find patterns, make decisions, and predict outcomes.

Supervised learning (SL) is a type of machine learning where the model is trained on labeled data, meaning each input has a corresponding known output. The goal is to learn a mapping from inputs to outputs to make accurate predictions on new, unseen data.

Unsupervised learning (UL) involves training a model on data without labeled outputs. The goal is to identify hidden patterns, groupings, or structures within the data.

2 Exercise 2

a) Classification is a learning task where the model predicts discrete labels or categories based on input data. It assigns data points to predefined classes. Real-life examples include email spam detection (spam or not spam), medical diagnosis (disease categories), and image recognition (identifying objects or people).

Regression is another learning task that predicts continuous numerical values. It estimates relationships between variables. Real-life examples include predicting house prices, forecasting stock prices, and estimating a person's income based on their education and experience.

b) Clustering is a learning technique that groups data points into clusters based on their similarities, without using labeled data. It helps discover hidden structures in the data. Examples include customer segmentation in marketing, grouping similar news articles, and anomaly detection in network security.

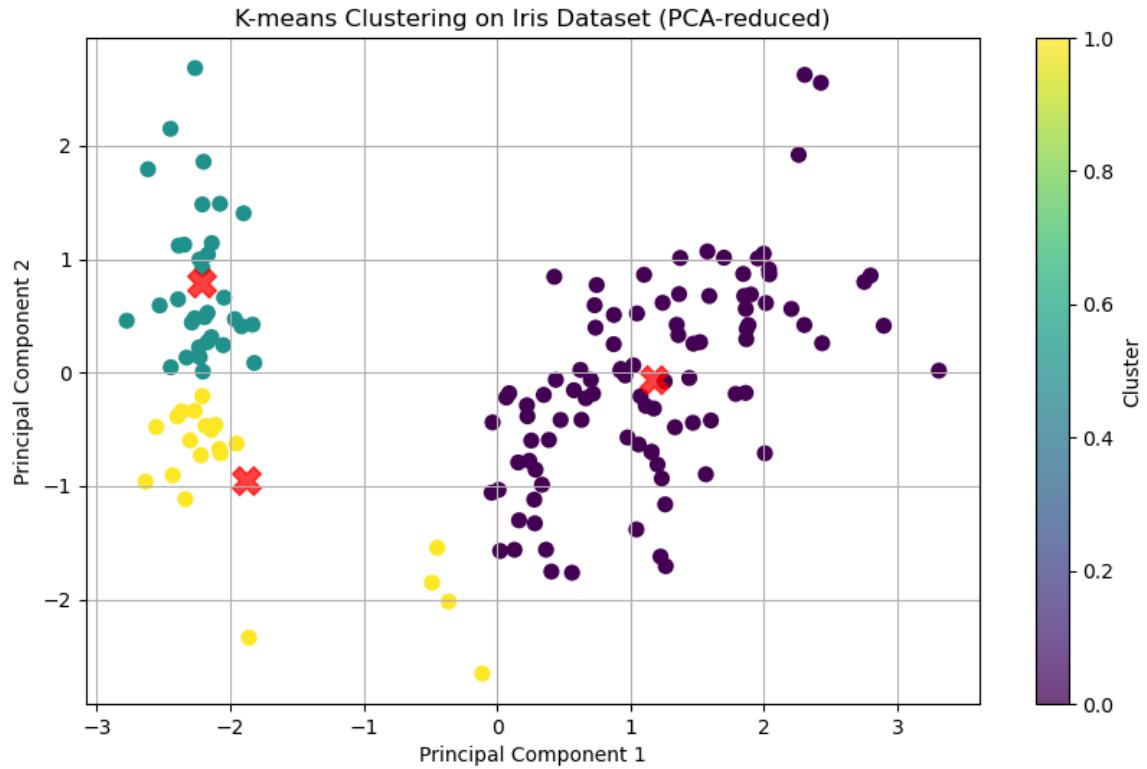
Dimensionality reduction is a technique used to reduce the number of features or variables in a dataset while preserving as much important information as possible. It simplifies complex data, making it easier to visualize and analyze. Examples include compressing image data, speeding up machine learning models, and visualizing high-dimensional data.

c) Classification and regression both belong to supervised learning because they rely on labeled data for training. In supervised learning, the model learns from input-output pairs, where the output is known. In classification, the model predicts discrete labels, learning from examples with known categories. In regression, the model predicts continuous values, learning from data with known numerical outcomes

Dimensionality reduction and clustering both belong to unsupervised learning because they work with data that lacks labeled outputs. Clustering groups data points based on their inherent similarities or patterns, without predefined categories. Since there are no labels guiding the grouping, the model identifies patterns autonomously. Dimensionality reduction reduces the number of features in the data while retaining key information. It helps simplify data or visualize it in lower dimensions, without needing labeled outcomes. Both techniques aim to uncover hidden structures or simplify the data rather than predict a specific output.

d) Reinforcement learning is a type of machine learning where an agent learns to make decisions by interacting with an environment. The agent receives feedback in the form of rewards or penalties based on its actions and learns to maximize the cumulative reward over time. Reinforcement learning focuses on learning a policy that maps situations to optimal actions. A real-life example of reinforcement learning is autonomous driving, where a car (the agent) learns to navigate roads, avoid obstacles, and follow traffic rules by receiving rewards for safe driving behaviors and penalties for collisions or rule violations.

3 Exercise 3



The plot above shows the results of applying K-means clustering to the Iris dataset, reduced to two dimensions using PCA. The data points are colored based on their assigned cluster, and the red "X" marks indicate the cluster centroids.