Machine Learning and Intelligent Systems

Unsupervised Learning - Setup

Maria A. Zuluaga

Jan 26, 2023

EURECOM - Data Science Department

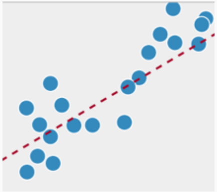
Table of contents

Recap

Supervised Learning: Learn with a teacher

Regression

Learning a continuous function from a set of examples

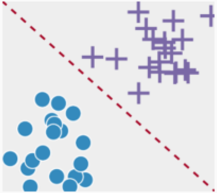


Examples:

- Forecasting
- Price of a good (stocks, flights, etc).
- Expected salary

Classification

Learning a rule from examples that can separate data from one another



Examples:

- Spam filtering
- Image classification
- Disease diagnosis

Unsupervised Learning

Unsupervised learning: No teacher

- No target or labels.
- The goal is to discover "interesting structure" in the data, often also referred to as knowledge discovery.
- Since there is no desired output, our task becomes one of **density estimation**.

Unsupervised learning: No teacher

- No target or labels.
- The goal is to discover "interesting structure" in the data, often also referred to as knowledge discovery.
- Since there is no desired output, our task becomes one of **density estimation**.

Formulation: The training set $\mathcal{D}$ consists of N observations $\mathbf{x}_i \in \mathbb{R}^D$ drawn from a probability distribution $p(X)$.

- **Supervised learning:** $p(\mathbf{y}|\mathbf{X})$
- **Unsupervised learning:** $p(\mathbf{X})$

Unsupervised learning: No teacher

- No target or labels.
- The goal is to discover "interesting structure" in the data, often also referred to as knowledge discovery.
- Since there is no desired output, our task becomes one of **density estimation**.

Formulation: The training set $\mathcal{D}$ consists of N observations $\mathbf{x}_i \in \mathbb{R}^D$ drawn from a probability distribution $p(X)$.

- **Supervised learning:** $p(\mathbf{y}|\mathbf{X})$
- **Unsupervised learning:** $p(\mathbf{X})$

Goal: Directly infer the properties of this probability density function (PDF) without the help of a supervisor providing correct answers.

Canonical Examples: Discovering Clusters

Goal: Attempts to find multiple convex regions of the X-space that contain modes of $p(\mathbf{X})$.

Allows to determine whether or not $p(\mathbf{X})$ can be represented by a mixture of simpler densities representing distinct types or classes of observations.

Canonical Examples: Discovering Clusters

Goal: Attempts to find multiple convex regions of the X-space that contain modes of $p(\mathbf{X})$.

Allows to determine whether or not $p(\mathbf{X})$ can be represented by a mixture of simpler densities representing distinct types or classes of observations.

Tasks:

- In the supervised case, we knew the expected number of clusters. In the unsupervised case, you decide this number (model selection).
- Determine which sample belongs to each cluster.

Canonical Examples: Discovering Clusters

Goal: Attempts to find multiple convex regions of the X-space that contain modes of $p(\mathbf{X})$.

Allows to determine whether or not $p(\mathbf{X})$ can be represented by a mixture of simpler densities representing distinct types or classes of observations.

Tasks:

- In the supervised case, we knew the expected number of clusters. In the unsupervised case, you decide this number (model selection).
- Determine which sample belongs to each cluster.

Real-world applications: Segmentation (images, groups of people, genes), sub-population of cells (biology), groups of stars (astronomy), novelty and outlier detection.

Canonical Examples: Discovering Clusters

Goal: Attempts to find multiple convex regions of the X-space that contain modes of $p(\mathbf{X})$.

Allows to determine whether or not $p(\mathbf{X})$ can be represented by a mixture of simpler densities representing distinct types or classes of observations.

Tasks:

- In the supervised case, we knew the expected number of clusters. In the unsupervised case, you decide this number (model selection).
- Determine which sample belongs to each cluster.

Real-world applications: Segmentation (images, groups of people, genes), sub-population of cells (biology), groups of stars (astronomy), novelty and outlier detection.

Algorithms: K-means, Hierarchical clustering, DBSCAN

Note: Mixture models have similar goals and are sometimes considered as clustering techniques

Canonical Examples: Dimensionality Reduction

Goal: Attempts to identify low-dimensional manifolds within the X-space that represent high data density.

Provides information about associations among variables and whether or not they can be considered as functions of a smaller set of latent variables.

The motivation behind this technique is that although the data may appear high dimensional, there may only be a small number of degrees of variability, corresponding to latent factors

Canonical Examples: Dimensionality Reduction

Goal: Attempts to identify low-dimensional manifolds within the X-space that represent high data density.

Provides information about associations among variables and whether or not they can be considered as functions of a smaller set of latent variables.

The motivation behind this technique is that although the data may appear high dimensional, there may only be a small number of degrees of variability, corresponding to latent factors

Real-world applications: interpretation of gene microarray data, data visualization, separation of signals into different sources (speech recognition).

Canonical Examples: Dimensionality Reduction

Goal: Attempts to identify low-dimensional manifolds within the X-space that represent high data density.

Provides information about associations among variables and whether or not they can be considered as functions of a smaller set of latent variables.

The motivation behind this technique is that although the data may appear high dimensional, there may only be a small number of degrees of variability, corresponding to latent factors

Real-world applications: interpretation of gene microarray data, data visualization, separation of signals into different sources (speech recognition).

Algorithms: Principal component analysis, self-organizing maps, autoencoders, principal curves, multidimensional scaling

Canonical Examples: Matrix Completion

Goal: Allows to complete data that may be missing from the collected data. This is often also called as imputation.

Real-world applications: Image inpainting, collaborative filtering, market basket analysis.

Algorithms: Non-negative matrix factorization

Canonical Examples: Matrix Completion

Goal: Allows to complete data that may be missing from the collected data. This is often also called as imputation.

Real-world applications: Image inpainting, collaborative filtering, market basket analysis.

Algorithms: Non-negative matrix factorization

Note: Scikit-learn proposes a very good catalog of unsupervised techniques, while using a different taxonomy [link] .

Wrap-up

Wrap-up

- We presented the setup for unsupervised learning
- We introduced some of the main techniques used

Key Concepts

- Probability density function
- Clustering
- Dimensionality reduction

References

Further Reading and Useful Material

Source	Notes
The Elements of Statistical Learning	Ch. 14