

Introduction to Machine Learning

Session 3a: Introduction to Unsupervised Learning

Reto Wüest

Department of Political Science and International Relations
University of Geneva

- ① Unsupervised Learning
- ② The Challenge of Unsupervised Learning

Unsupervised Learning

- Recall that in an **unsupervised learning** problem we only have a set of features X_1, X_2, \dots, X_p measured on n observations.
- Our focus is on **two types** of unsupervised learning techniques:
 - **Principal components analysis:** Used for data visualization or data pre-processing before supervised learning techniques are applied;
 - **Clustering methods:** Used for discovering unknown subgroups in the data.

The Challenge of Unsupervised Learning

The Challenge of Unsupervised Learning

- Unsupervised learning is often **more challenging** than supervised learning:
 - It is more **subjective**;
 - It is more difficult to **assess the results**.
- If we use a supervised learning technique to predict Y , then we can **check our work** by seeing how well our model predicts Y on observations in a test set.
- In contrast, in unsupervised learning, we **cannot check our work** because we do not know the true answer (the problem is unsupervised).