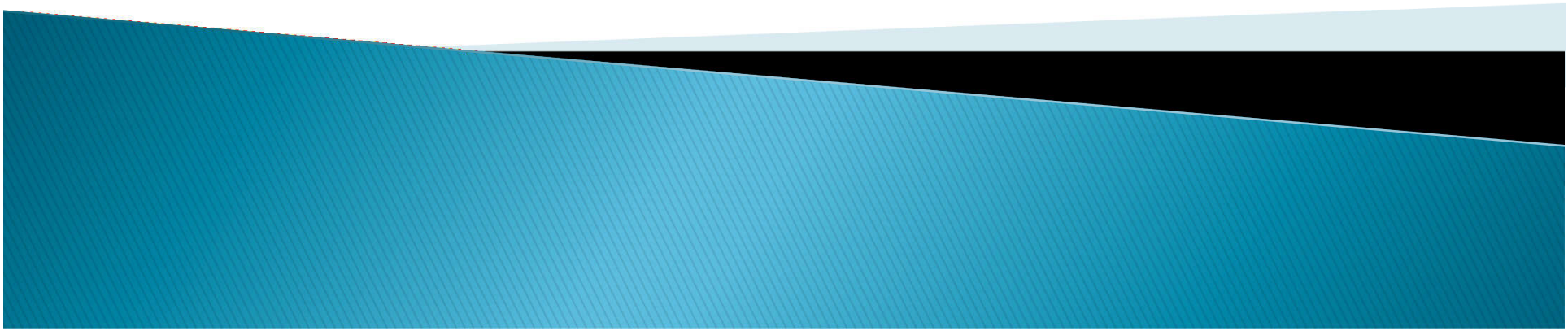
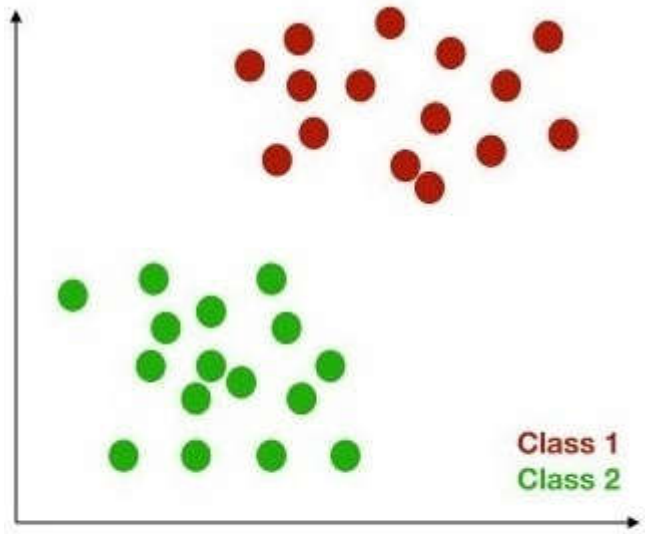
Unsupervised Learning

Nazerfard, Ehsan
nazerfard@aut.ac.ir

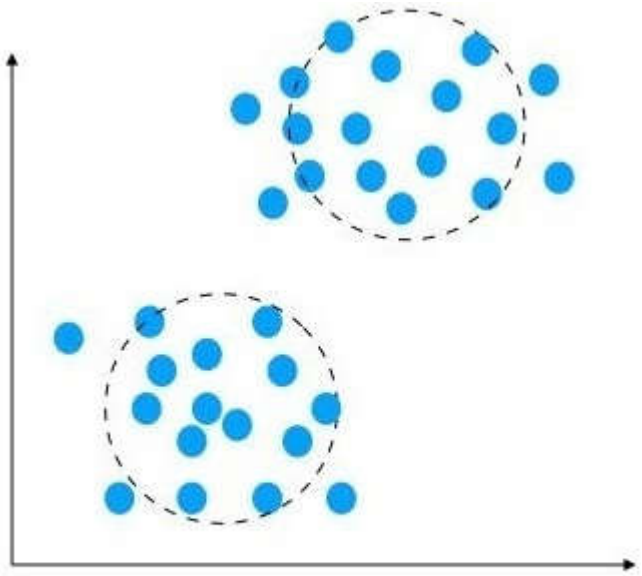


Supervised vs. Unsupervised

- Predictive Modeling
- Descriptive Modeling



Supervised



Unsupervised

$$D1_{train} = \{(\boldsymbol{X_1}, y_1), (\boldsymbol{X_2}, y_2) \ldots, (\boldsymbol{X_m}, y_m)\}$$

$$D2_{train} = \{\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_m}\}$$

Unsupervised Learning

- In supervised learning, we are given features together with targets (i.e. class labels).
- In unsupervised learning, we are only given features.
  - The task is to discover structures in the data.
- Why discover structures?

Unsupervised Learning (cont.)

- In supervised learning, we are given features together with targets (i.e. class labels).
- In unsupervised learning, we are only given features.
  - The task is to discover structures in the data.
- Why discover structures?
  - More meaningful representation for the data
    - Dimensionality reduction
    - Density Estimation
    - Clustering
    - Pre-processing technique prior to applying another ML model
    - …

Dimensionality Reduction

- Discover a lower-dimensional surface on which the data lives.

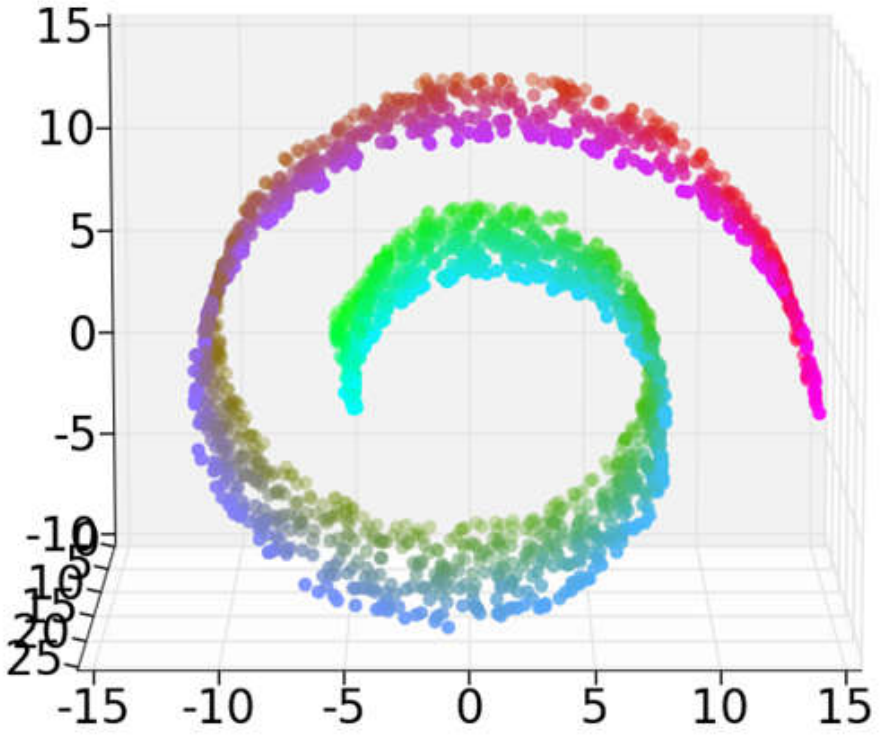


Figure © C. Schmid and J. Verbeek, Lecture on Machine Learning and Category Recognition, INRIA-LEAR

Density Estimation

- Find a function that approximates the probability density of the data (i.e., value of the function is high for "typical" points and low for "atypical" points).

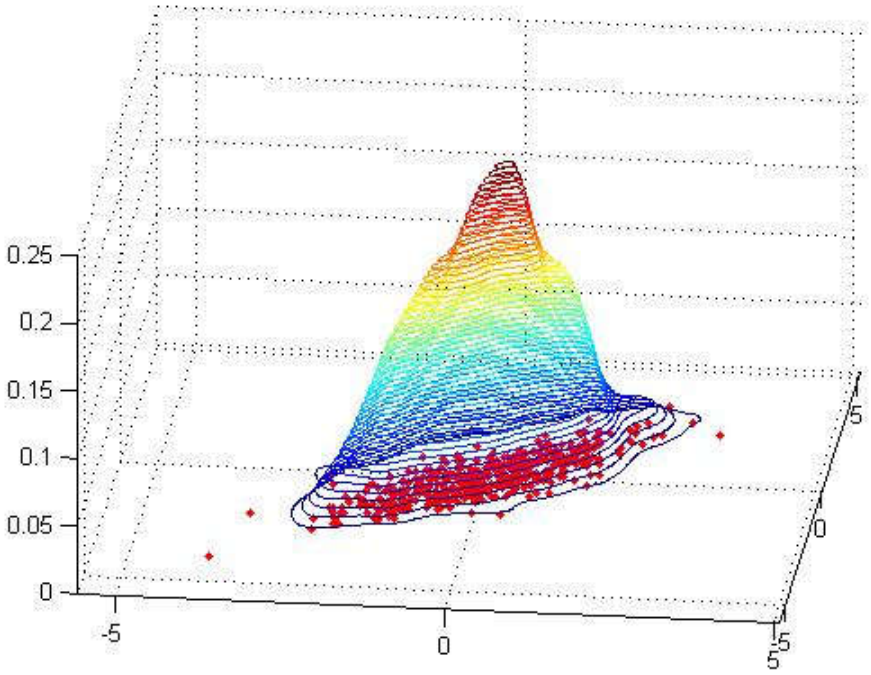


Figure © C. Schmid and J. Verbeek, Lecture on Machine Learning and Category Recognition, INRIA-LEAR

Clustering

- **Clustering** is the task of grouping a set of objects in such a way that objects in the same group (called a **cluster**) are more similar (in some sense) to each other than to those in other groups.

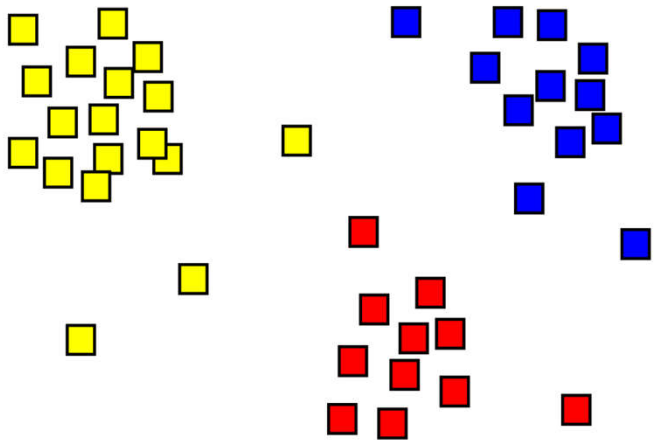


Figure © Wikipedia

Pre-processing

- Clustering can be used in process of semi-supervised Learning. It can be used first to find natural segmentation of the data and then create labels.
- The pre-processed data with labels can then be used to develop a semi-supervised classification.

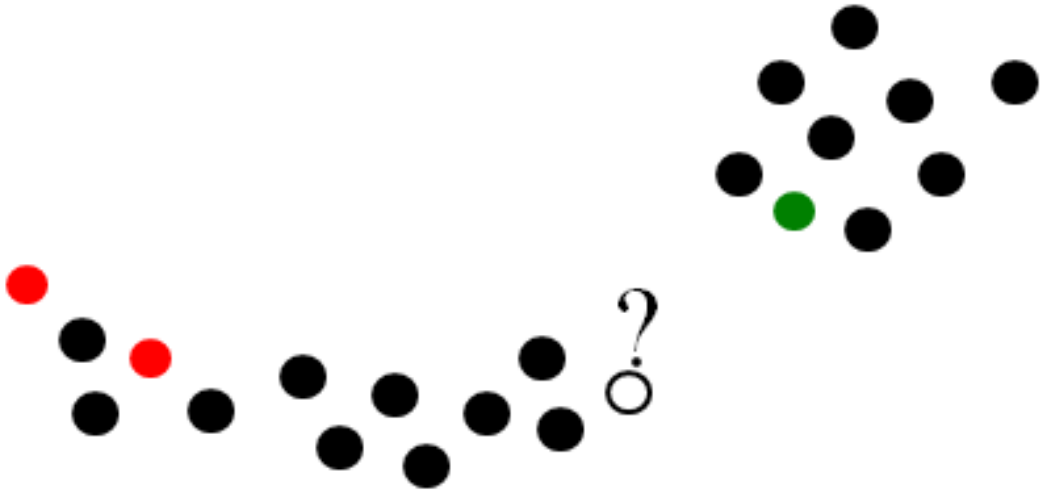


Figure © C. Schmid and J. Verbeek, Lecture on Machine Learning and Category Recognition, INRIA-LEAR

Further Reading (cont.)

- Semi-supervised learning
  - Lots of data is available, but only small portion is labeled.
  - J van Engelen, H Hoos, A survey on semi-supervised learning, Machine Learning, 109:373-440, 2020.

- R Raina, A Battle, H Lee, B Packer, A Ng, Self-taught learning: transfer learning from unlabeled data, ICML '07: Proceedings of the 24th international conference on Machine learning, pp 759-766, 2007.