Unsupervised Learning & Data Clustering

Ira A. Fulton Schools of Engineering
Arizona State University

Unsupervised Learning – Part 1: Problem Set-up

Ira A. Fulton Schools of Engineering
Arizona State University

Objective

Objective

Define the set-up of unsupervised learning

Learning from Unlabeled Data

- Given a training set of n **unlabeled** samples $\{\mathbf{x}^{(i)}\}$
- What can we learn from the samples?

➔ We could estimate the overall distribution of the data without knowing their label.

➔ We could figure out the groupings of the samples (if any).

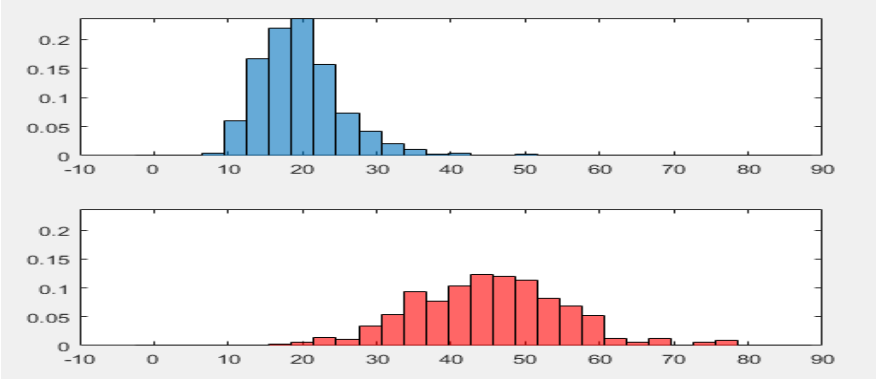
➔ We could identify some features that may be more important than others.

…

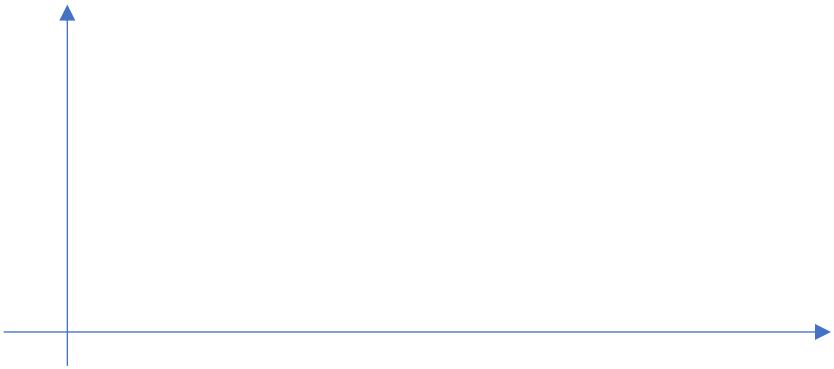
An Example

Illustrating structures/groupings of unlabeled samples may relate to the (unknown) labels of the samples

➔ If we know the labels, we may find the densities of the classes →

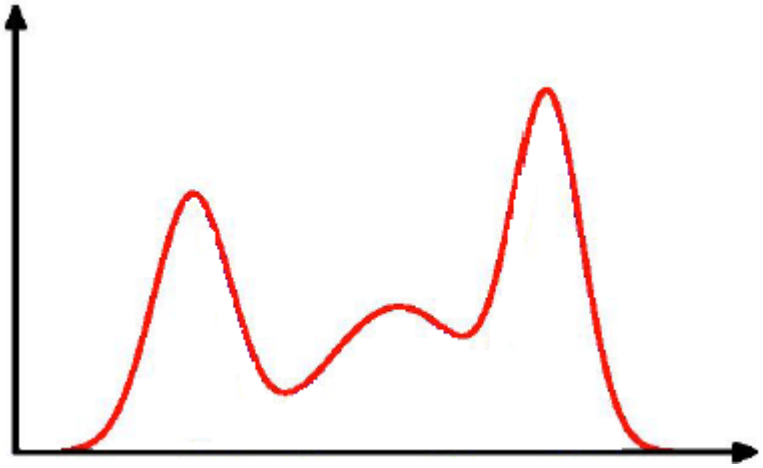


➔ What may we see if we have no label for the data samples?



Another Example

- A density estimated from unlabeled samples may help us to identify densities of different classes



- If we know there are three classes in the data, each having a normal distribution …

A Mixture-Density Model

Assume a parametric model like this:

- The samples come from C classes.
- The prior probabilities $P(\omega_j)$ for each class are known, for $j = 1, \ldots, C$.
- The form of $p(\mathbf{x} \mid \omega_j, \theta_j)$ $(j = 1, \ldots, C)$ are known.
- The C parameter vectors $\theta_1, \theta_2, \ldots, \theta_C$ are unknown.

Samples from this distribution are given, but the labels of the training samples are *unknown*.

A Mixture-Density Model (cont'd)

- What is the PDF of the unlabeled samples?

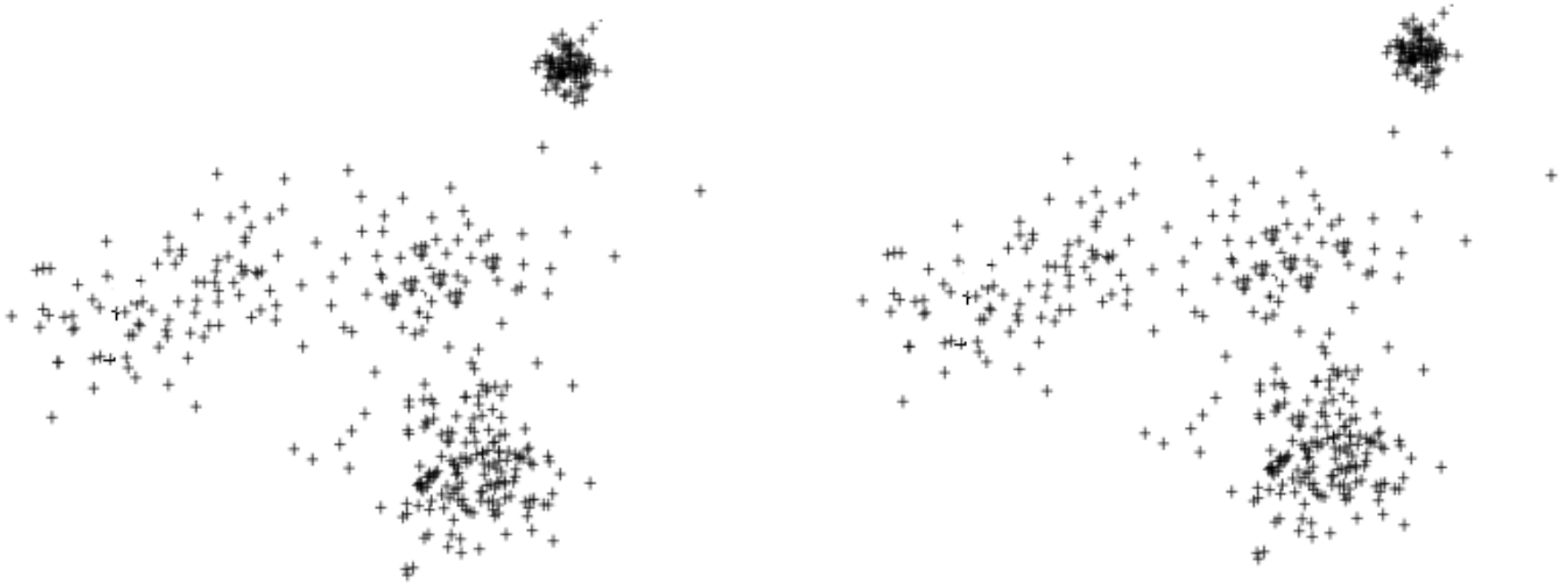
$$p(\mathbf{x}\,|\,\boldsymbol{\theta}) = \sum_{j=1}^{C} p(\mathbf{x}\,|\,\omega_j, \boldsymbol{\theta}_j) P(\omega_j)$$

where $\theta = (\theta_1, \theta_2, \ldots, \theta_C)$

- Can we learn θ from unlabeled samples from this **mixture density**?

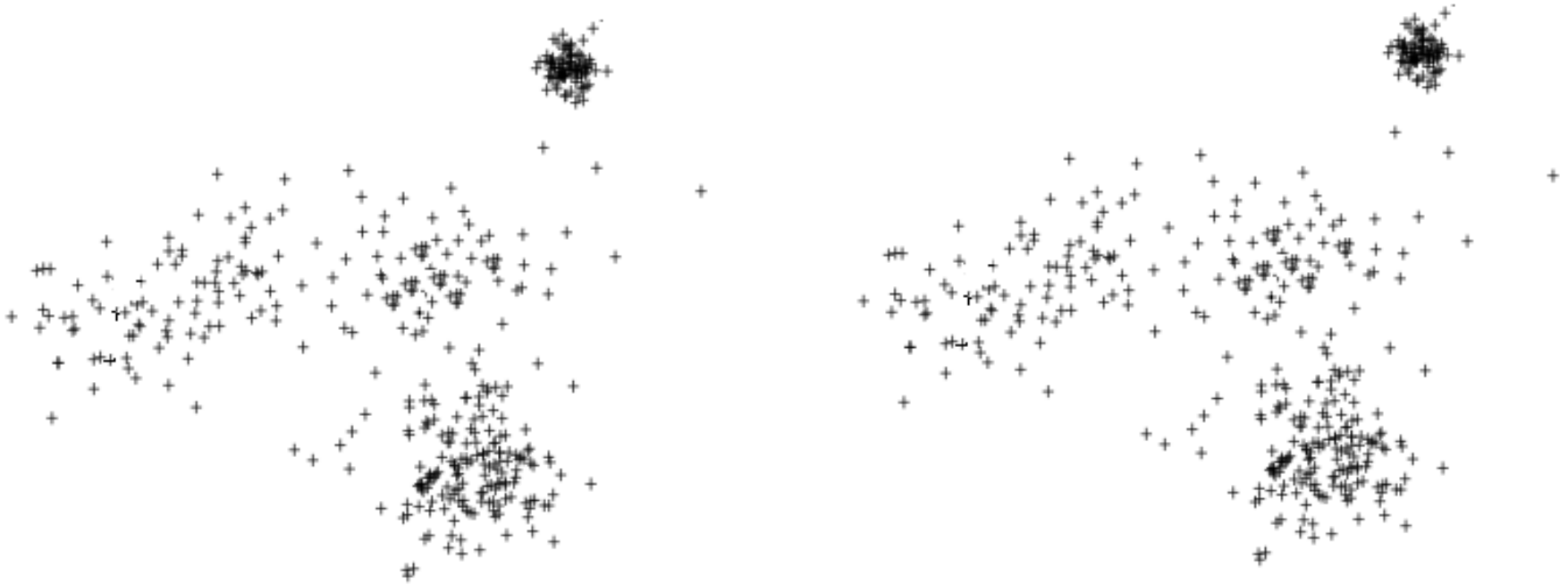
Illustrating Mixture-Density Model

- An example: with the assumption of 4 classes



Illustrating Mixture-Density Model

- An example: with the assumption of 2 classes



The Question of Identifiability

- Can we learn a unique θ from a set of unlabeled samples from a mixture density?
  - For continuous features (with PDFs), the answer is often “Yes”.
- An example in discrete case (with PMF).
  - Two coins with P(head) being p & q respectively.
  - Randomly pick one and toss it; Record the outcome.
  - With only the outcomes of N tosses, but not knowing which coin was used each time (→ unsupervised), can we figure out p and q?