Unsupervised Learning – Part 2: Gaussian Mixture Models and the EM Algorithm

Ira A. Fulton Schools of Engineering
Arizona State University

Objective

Objective

Define the Gaussian Mixture Model

Objective

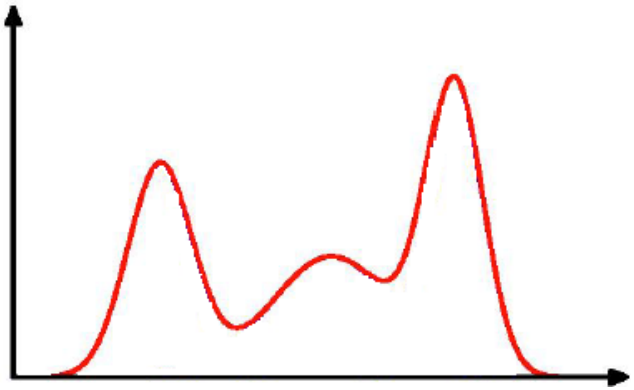
Illustrate the Expectation-Maximization Algorithm

The Gaussian Mixture Models

- The mixture model:

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

- GMM: each component density is a Gaussian distribution.
  - Can be a good approximation to many real data distributions.



If we do have labels …

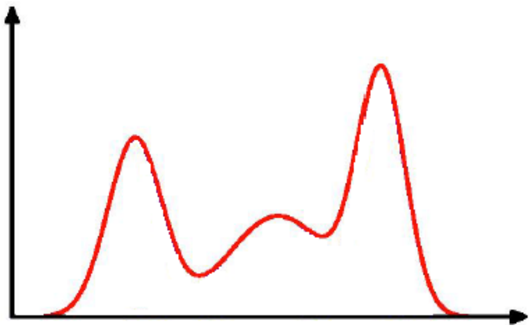
$$p(\mathbf{x} \mid \boldsymbol{\theta}) = \sum_{j=1}^{C} p(\mathbf{x} \mid \omega_j, \boldsymbol{\theta}_j) P(\omega_j)$$

This becomes supervised learning for each component (class).

It is more difficult without labels.

Unsupervised Case

- Consider an iterative method using the *maximum likelihood estimation* concept.



- Consider a 3-component 1-d example.

- What are the parameters in this case?

- We might have some initial (imprecise) guesses for the parameter, e.g., vs the *k-means algorithm*.
  - *How to improve the initial guesses?*

Unsupervised Case (cont'd)

An example of Expectation-Maximization Algorithm.

Iterate on t

Given parameter estimates at iteration t-1

Step 1. For a sample j, compute its probability of being from class *k*

$$P[y_j = k \mid x_j, \theta^{(t-1)}] \propto p_k^{(t-1)} \, p(x_j \mid \mu_k^{(t-1)}, \sigma_k^{(t-1)}), \; \forall k = 1, 2, 3$$

Step 2. Update the estimates of the parameters

$$\mu_k^{(t)} = \frac{\sum_j x_j P[y_j = k \mid x_j, \theta^{(t-1)}]}{\sum_j P[y_j = k \mid x_j, \theta^{(t-1)}]}$$

$$p_k^{(t)} = \frac{\sum_j P[y_j = k \mid x_j, \theta^{(t-1)}]}{\text{total \# of samples}}$$