Mixture Models

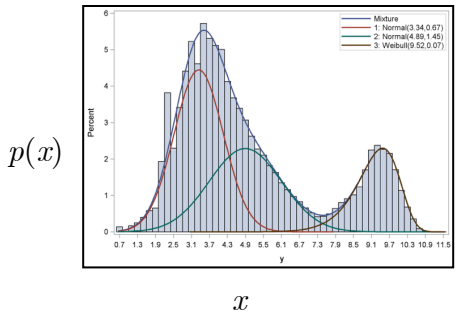
Sargur Srihari
srihari@cedar.buffalo.edu

Topics in Mixture Models and EM

- Modeling complex distributions
- K-means algorithm for finding clusters in a data set
- Latent variable view of mixture distributions
- General technique for finding m.l. estimators in latent variable models
- EM Algorithm
- Infinite Mixture Models

Modeling complex distributions

- Complex distribution $p(x)$ of observed variable x

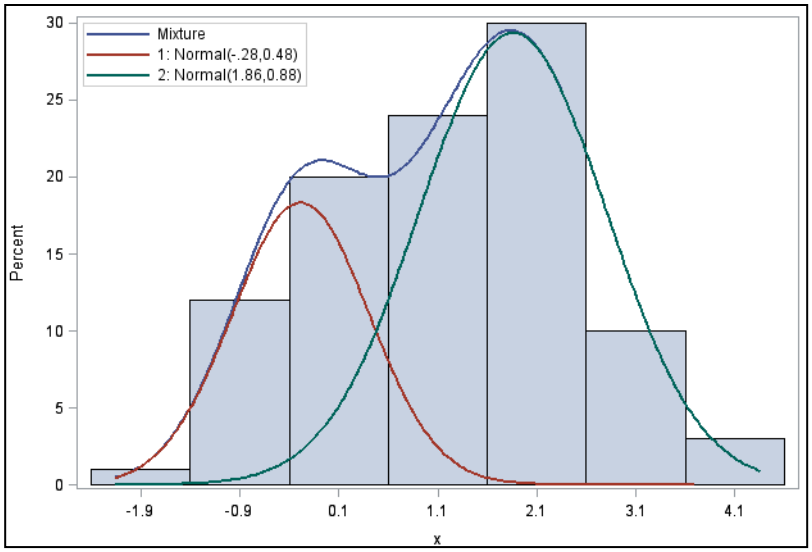


  - Can be expressed in terms of a more tractable joint distribution over observed and latent variables $p(x,z)$
    - Latent variable z with three values can model this distribution
  - Distribution of x alone obtained by marginalization $\boxed{p(x)=\sum_z p(x,z)}$
- Latent variables allow complicated distributions to be formed from simpler components
  - Gaussian mixtures have latent variables z that are discrete
  - Also called Finite Mixture Models

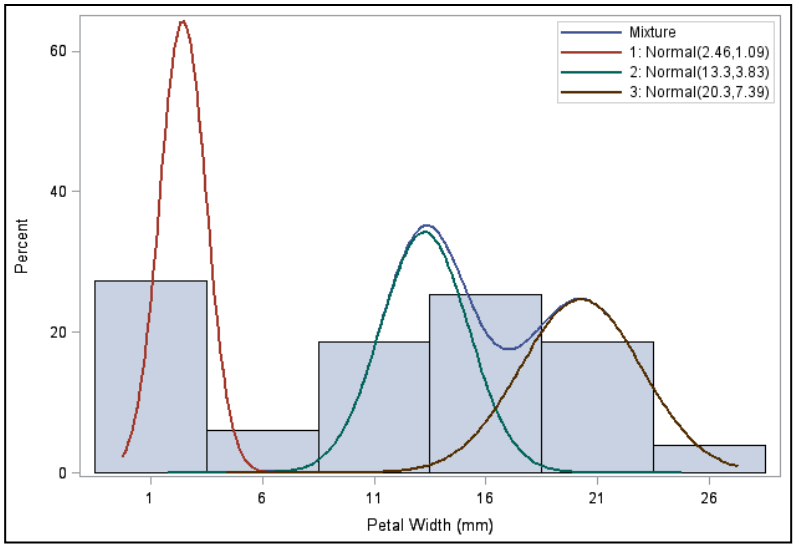
Gaussian Mixture Model (GMM)

- Linear superposition of Gaussian components
  - Two Gaussians
  - Three Gaussians

z has 2 values



z has 3 values (data: petal width in Iris)



Since $$p(x) = \sum_z p(x,z) = \sum_z p(z)p(x \mid z)$$

We can write (for the mixture of two Gaussians): $$p(x)=p(z=1)p(x|z=1)+p(z=2)p(x|z=2)$$

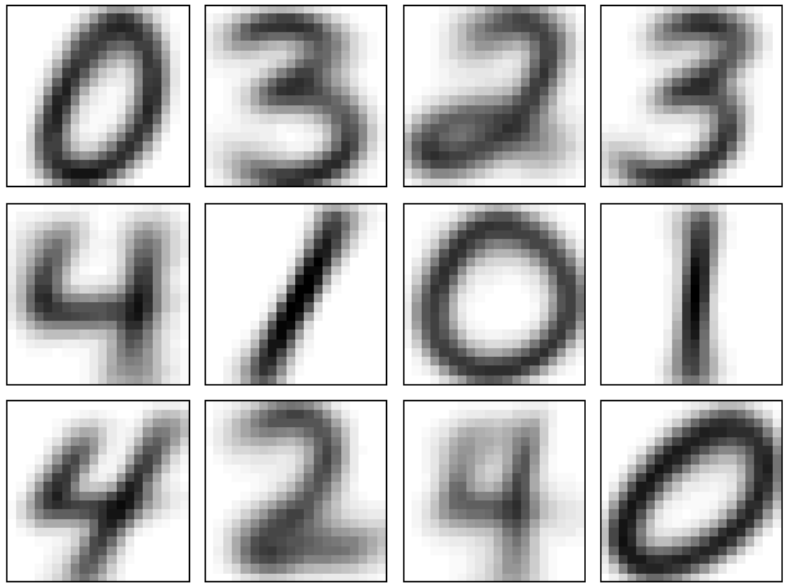
Mixture Model As Unsupervised Learning

- Probabilistic model representing sub-populations within a population
  - Without requiring that the sub-population of the data items be identified (supervised)
- Constructing such models is called unsupervised learning or clustering

Bernoulli Mixture Model

- Handwritten Digit Data (560×420 pixels)
  - Mixture Model for digits 0-4 with $K = 12$
    - Identifies three 0s, two 1s,two 3s, and three 4s

Superimposed data of 12 components



Role of Mixture Models

- Mixture models provide:
  1. Framework for building complex probability distributions
     - Complex distribution expressed in terms of tractable joint distribution of observed and latent variables
       - Distribution of observed variables: by marginalization
  2. A method for clustering data
     - Unsupervised learning