Foundations of Machine Learning
AI2000 and AI5000

FoML-28
Latente Variable Models, GMM, and EM

Dr. Konda Reddy Mopuri
Department of AI, IIT Hyderabad
July-Nov 2025

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

DiL
Data-driven Intelligence
& Learning Lab

So far in FoML

- Intro to ML and Probability refresher
- MLE, MAP, and fully Bayesian treatment
- Supervised learning
    a. Linear Regression with basis functions
    b. Bias-Variance Decomposition
    c. Decision Theory - three broad classification strategies
    d. Neural Networks
- Unsupervised learning
    a. K-Means, Hierarchical, and GMM for clustering

For today

- Latent Variable Models

Supervised vs. Unsupervised learning

- Data {X, T} is given
- Goal: mapping f(x) ≅ t

Reg $p(t/x)$

Clasf $p(t/x)$

predictive distribution

or

point prediction

t

- Data {X} is given
- Goal: interesting aspects of data

- density $p(x)$ estimation
- structure (clusters)
- Dimensionality reduction

Latent Variable Models

- Model complex distributions with more tractable representation
  - Via z

Latent variable
(unobserved)

Observed variable

Latent Variable Models

- Model complex distributions with more tractable representation
  - Via z
- Continuous latent variable ‘z’

$$p(x) = \int p(x,z)\,dz = \int p(x|z)\,p(z)\,dz$$

Latent variable (unobserved)

Observed variable

Latent Variable Models

- Model complex distributions with more tractable representation
  - Via z
- Discrete latent variable 'z'

$$p(x) = \sum_{z} p(x, z) = \sum_{z} p(x|z)\, p(z)$$

Latent variable (unobserved)

Observed variable

GMM in terms of discrete latent variables

- Gaussian mixture distribution can be written as a superposition of multiple Gaussians

$$p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\underline{x}, \underline{\mu_k}, \Sigma_k)$$

z { cluster index / class/category

$\underline{x}$ data

GMM in terms of discrete latent variables

- Let's introduce a K-dim binary random variable 'z'
    - 1-of-K representation

$\underline{Z} = \{z_1\ z_2\ \cdots\ z_K\}$

$z_i = \{0, 1\}$

$\sum_{i=1}^{K} z_i = 1$

GMM in terms of discrete latent variables

- We shall define the joint distribution in terms of the conditional and marginal

$$p(\mathbf{x}, \mathbf{z}) = \quad P(X/z) \; P(z)$$

GMM in terms of discrete latent variables

- The marginal over the latent variable is expressed in terms of the mixing coefficients

$$p(\mathbf{z}_k = 1) = \pi_k$$

$$p(\mathbf{z}) = \prod_{i=1}^{K} (\pi_i)^{z_i}$$

(z)

$z_i = 1$

when $\underline{x} \sim$ ith component of the mixture

GMM in terms of discrete latent variables

- The conditional distribution of x given a particular value of z is a Gaussian

$$p(\mathbf{x}|z_k = 1) = \mathcal{N}(\underline{x}, \underline{\mu}_k, \Sigma_k)$$

$$p(\mathbf{x}|\mathbf{z}) = \prod_{i=1}^{K} \mathcal{N}(\underline{x}, \underline{\mu}_i, \Sigma_i)^{z_i}$$

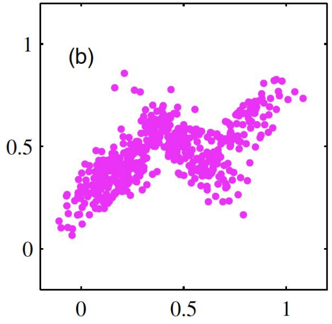
$$p(\underline{x}) = \sum_{\underline{z}} p(\underline{z})\, p(\underline{x}|\underline{z}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\underline{x}, \underline{\mu}_k, \Sigma_k)$$

GMM in terms of discrete latent variables

- Conditional probability of z given x

$$\gamma(z_k) \equiv p(z_k = 1|\mathbf{x}) = \frac{P(X|z_k=1)\; P(z_k=1)}{\sum_{i=1}^{K} P(X|z_i=1)\; P(z_i=1)}$$

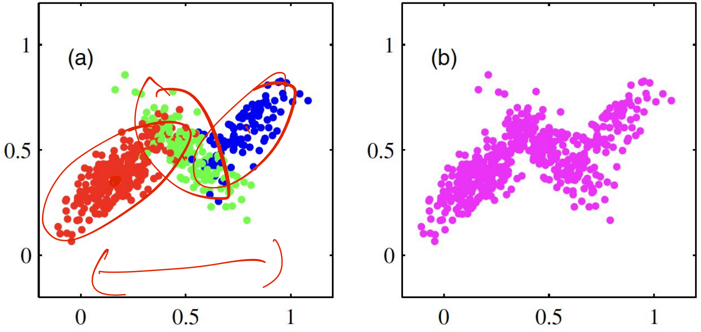
GMM example



- 500 samples from marginal p(x)

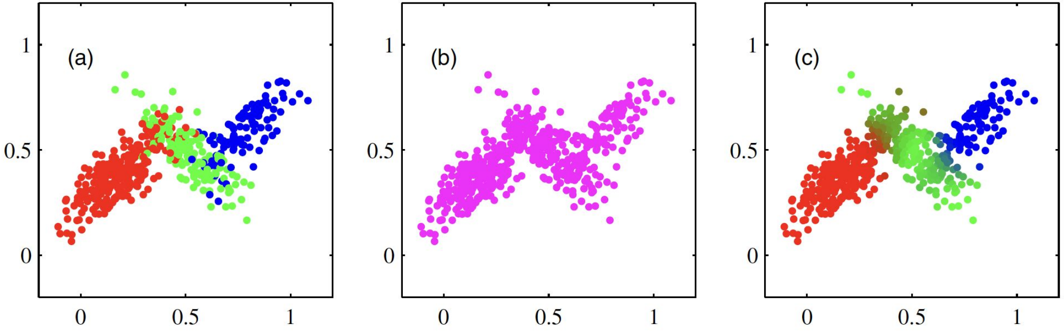
without revealing the latent variable

GMM example



- [Left] same samples drawn from the joint of distribution ρ(x/z) ρ(z)
  - Complete dataset (doesn't ignore the latent variable)

GMM example



$p(z_{red}=1|x)$

$p(z_{green}=1|x)$

$p(z_{blue}=1|x)$

- [Right] same samples with colors representing the $\gamma(z_k)$

Modeling the data with GMM

Modeling using GMMs

- Data of iid observations $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$

- The Log-likelihood is given by

$$\ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \sum_{i=1}^{N} \ln\left\{ \sum_{k=1}^{K} \pi_k \cdot \mathcal{N}(\underline{x_i} \mid \underline{\mu_k}, \Sigma_k) \right\}$$

Modeling using GMMs

- Setting the derivatives of the Log-likelihood gives

$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{x}_n$$

$$N_k = \sum_{n=1}^{N} \gamma(z_{nk})$$

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) (\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^{\mathrm{T}}$$

Constrained ⇒ Lagrange multiplier

$$\pi_k = \frac{N_k}{N}$$

$$\sum_{i=1}^{K} \pi_i = 1$$

$$\ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) + \lambda \left(\sum_{k=1}^{K} \pi_k - 1 \right)$$

Modeling using GMMs

- Setting the derivatives of the Log-likelihood gives

$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{x}_n$$

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) (\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^{\mathrm{T}}$$

$$\pi_k = \frac{N_k}{N}$$

Note that these are not 'closed from' solutions because of the dependency of $\gamma(z_k)$

$$\gamma(z_k) \equiv p(z_k = 1|\mathbf{x}) = \frac{p(z_k = 1)p(\mathbf{x}|z_k = 1)}{\sum_{j=1}^{K} p(z_j = 1)p(\mathbf{x}|z_j = 1)} = \frac{\pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}.$$

That is why we take an iterative approach!

EM algorithm for GMM

EM for Gaussian Mixtures

Given a Gaussian mixture model, the goal is to maximize the likelihood function with respect to the parameters (comprising the means and covariances of the components and the mixing coefficients).

1. Initialize the means $\boldsymbol{\mu}_k$, covariances $\boldsymbol{\Sigma}_k$ and mixing coefficients π_k, and evaluate the initial value of the log likelihood.
2. **E step**. Evaluate the responsibilities using the current parameter values

$$\gamma(z_{nk}) = \frac{\pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}. \tag{9.23}$$

3. **M step**. Re-estimate the parameters using the current responsibilities

$$\boldsymbol{\mu}_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{x}_n \tag{9.24}$$

$$\boldsymbol{\Sigma}_k^{\text{new}} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \left(\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}}\right) \left(\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{new}}\right)^{\mathrm{T}} \tag{9.25}$$

$$\pi_k^{\text{new}} = \frac{N_k}{N} \tag{9.26}$$

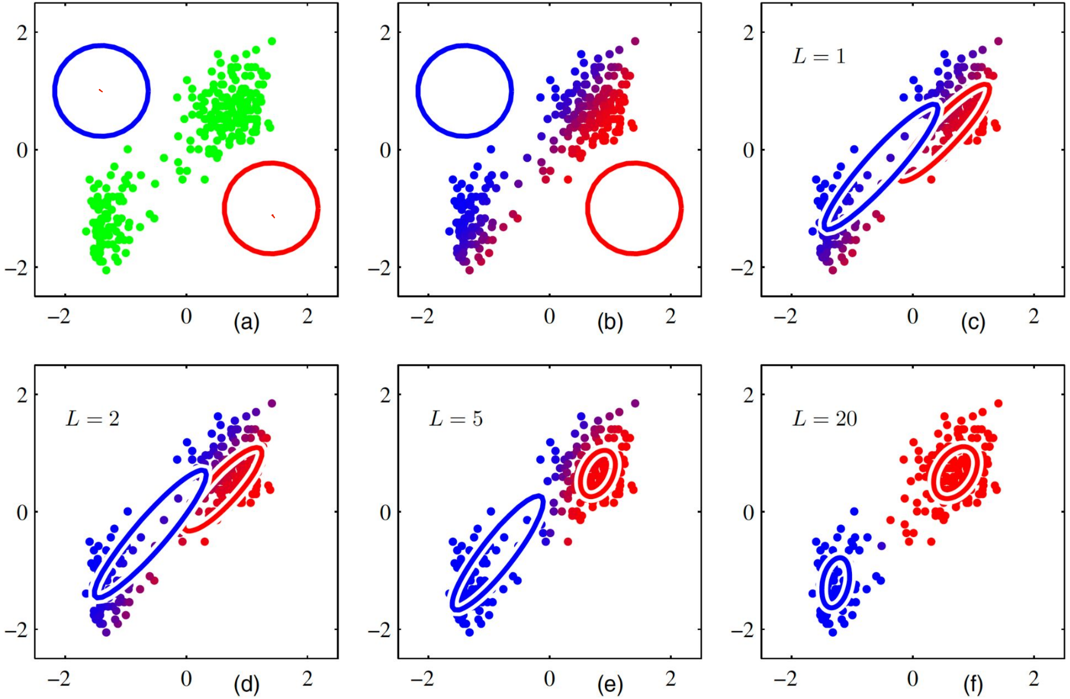
where

$$N_k = \sum_{n=1}^{N} \gamma(z_{nk}). \tag{9.27}$$

4. Evaluate the log likelihood

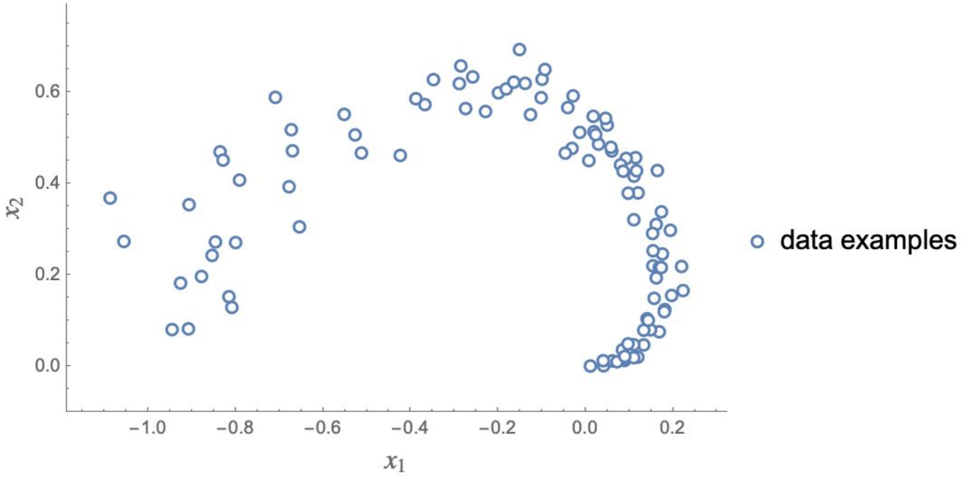
$$\ln p(\mathbf{X}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\} \tag{9.28}$$

EM algorithm illustration



Latent variables - connection to dimensionality reduction

Manifold coordinates as Latent variables



$$\{x_1, x_2\} = \{t\cos(3\,t),\ t\sin(3\,t)\}$$

t

Figure credits

Next

- PCA