Expectation Maximization
Chapter 5
Gaussian Mixture Models II

T.L. Grobler

Supervised versus Unsupervised

- Supervised Learning:
  - Chapter 4
  - Class label comes with observations
- Unsupervised Learning:
  - Chapter 5
  - Class label inferred from observations

Expectation Maximization

- Expectation:
  - Chapter 5
  - Current estimate of parameters
  - Estimate class labels
- Maximization (MLE):
  - Chapter 4
  - Current estimate of class labels
  - Estimate parameters
- Alternate E and M step till convergence

Introduction

- Single Gaussian pdf
  - Useful analytic properties
  - Not always practical
  - Multi-modal data set
  - Mixture of *k* Gaussians

A Gaussian Mixture

$$p(\mathbf{x}|\theta) = \sum_{j=1}^{k} \pi_j \mathcal{N}(\mathbf{x}|\mathbf{u}_j, \mathbf{\Sigma}_j)$$

- $\mathcal{N}$ – Gaussian distribution
- $\mathbf{u}$, $\mathbf{\Sigma}$ – Mean and Covariance
- $\boldsymbol{\pi}$ – Mixture coefficients (class probabilities)
- $\Sigma\pi=1$
- $\boldsymbol{\theta}$ – Mixture Parameters ($\mathbf{u}$, $\mathbf{\Sigma}$, $\boldsymbol{\pi}$)

Problem

- Estimate $\boldsymbol{\theta}$ from $\mathbf{x}$
  - Contribution to mixture components unknown
- Similar to *k*-means
  - Cluster label unkown

A General Mixture Model

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

Marginalization

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

Chain-Rule

Enumeration

$$= \sum_{j=1}^{k} \underbrace{p(\mathbf{x}|z_j = 1)}_{\mathcal{N}}\underbrace{P(z_j = 1)}_{\pi_j}$$

Gaussian Ass.

$$\mathbf{z} = [z_1, \ \cdots \ , z_n]^T \text{ where } z_j \in \{0, 1\} \text{ and } \sum_j z_j = 1$$

Latent-variable

Mixture Component Prob γ_j

$$\gamma(z_j) = P(z_j = 1|\mathbf{x})$$

$$= \frac{P(z_j = 1)p(\mathbf{x}|z_j = 1)}{\sum_{i=1}^{k} P(z_i = 1)p(\mathbf{x}|z_i = 1)}$$

$$= \frac{\pi_j \mathcal{N}(\mathbf{x}|\mathbf{u}_j, \mathbf{\Sigma}_j)}{\sum_{i=1}^{k} \pi_i \mathcal{N}(\mathbf{x}|\mathbf{u}_i, \mathbf{\Sigma}_i)}$$

Bayes-law

Gaussian Ass.

Log-likelihood

Matrix *Nxd*

Chain-Rule

$$p(\mathbf{X}, \pi, \mathbf{u}, \Sigma) = p(\mathbf{X}|\pi, \mathbf{u}, \Sigma)p(\boldsymbol{\pi}, \boldsymbol{u}, \Sigma)$$

Independence

$$= p(\boldsymbol{\pi}, \boldsymbol{u}, \Sigma) \prod_{n=1}^{N} p(\mathbf{x}_n|\boldsymbol{\pi}, \boldsymbol{u}, \Sigma)$$

Substitution

$$= p(\boldsymbol{\pi}, \boldsymbol{u}, \Sigma) \prod_{n=1}^{N} \sum_{i=1}^{k} \pi_i \mathcal{N}(\mathbf{x}_n|\mathbf{u}_i, \boldsymbol{\Sigma}_i)$$

Discard

$$\ln p(\mathbf{X}|\pi, \mathbf{u}, \Sigma) = \sum_{n=1}^{N} \ln \left\{ \sum_{i=1}^{k} \pi_i \mathcal{N}(\mathbf{x}_n|\mathbf{u}_i, \boldsymbol{\Sigma}_i) \right\}$$

ML Estimation - mean

$$0 = -\sum_{n=1}^{N} \frac{\pi_j \mathcal{N}(\mathbf{x}_n | \mathbf{u}_j, \mathbf{\Sigma}_j)}{\sum_i \pi_i \mathcal{N}(\mathbf{x}_n | \mathbf{u}_i, \mathbf{\Sigma}_i)} \mathbf{\Sigma}_j^{-1} (\mathbf{x}_n - \mathbf{u}_j)$$

$$= -\sum_{n=1}^{N} \gamma(z_{nj}) \mathbf{\Sigma}_j^{-1} (\mathbf{x}_n - \mathbf{u}_j)$$

Chain Rule

$$\mathbf{u}_j = \frac{1}{N_j} \sum_{n=1}^{N} \gamma(z_{nj}) \mathbf{x}_n$$

$$\frac{d \ln x}{dx} = x^{-1}; \quad \frac{d e^{x^2}}{dx} = 2x e^{x^2}$$

$$N_j = \sum_{n=1}^{N} \gamma(z_{nj})$$

ML Estimation - covariance

$$\mathbf{\Sigma}_j = \frac{1}{N_j} = \sum_{n=1}^{N} \gamma(z_{nj})(\mathbf{x}_n - \mathbf{u}_j)(\mathbf{x}_n - \mathbf{u}_j)^T$$

ML Estimation – mixture coeff

Lagrange

$$\sum \ln\left\{\sum_{i=1}^{k} \pi_i \mathcal{N}(\mathbf{x}_n|\mathbf{u}_i, \boldsymbol{\Sigma}_i)\right\} + \lambda\left(\sum_{i=1}^{k}\pi_i - 1\right)$$

max

$$0 = -\sum_{n=1}^{N} \frac{\mathcal{N}(\mathbf{x}_n|\mathbf{u}_j, \boldsymbol{\Sigma}_j)}{\sum_i \pi_i \mathcal{N}(\mathbf{x}_n|\mathbf{u}_i, \boldsymbol{\Sigma}_i)} + \lambda$$

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

$\times \pi_j$ and $\sum_j$

$\times \pi_j$ only

+substitution of λ

$$\lambda = -N$$

$$0 = \sum_{n=1}^{N} \gamma(z_{nj}) - \pi_j N$$

$$\boxed{\pi_j = \frac{N_j}{N}}$$

EM Algorithm for GMMs

1) Initialize parameters $\boldsymbol{\pi}$, $\mathbf{u}$, $\boldsymbol{\Sigma}$ (k-means).

2) **E-step**: Estimate responsibilities λ

3) **M-step**: Re-estimate parameters $\boldsymbol{\pi}$, $\mathbf{u}$, $\boldsymbol{\Sigma}$

4) Check for convergence of parameters or log-likelihood (increase monotonically)

5) Iterate from (2) until convergence

E-step

$$\gamma(z_{nj}) = \frac{\pi_j \mathcal{N}(\mathbf{x}_n | \mathbf{u}_j \mathbf{\Sigma}_j)}{\sum_{i=1}^{k} \pi_i \mathcal{N}(\mathbf{x}_n | \mathbf{u}_i \mathbf{\Sigma}_i)}$$

M-step

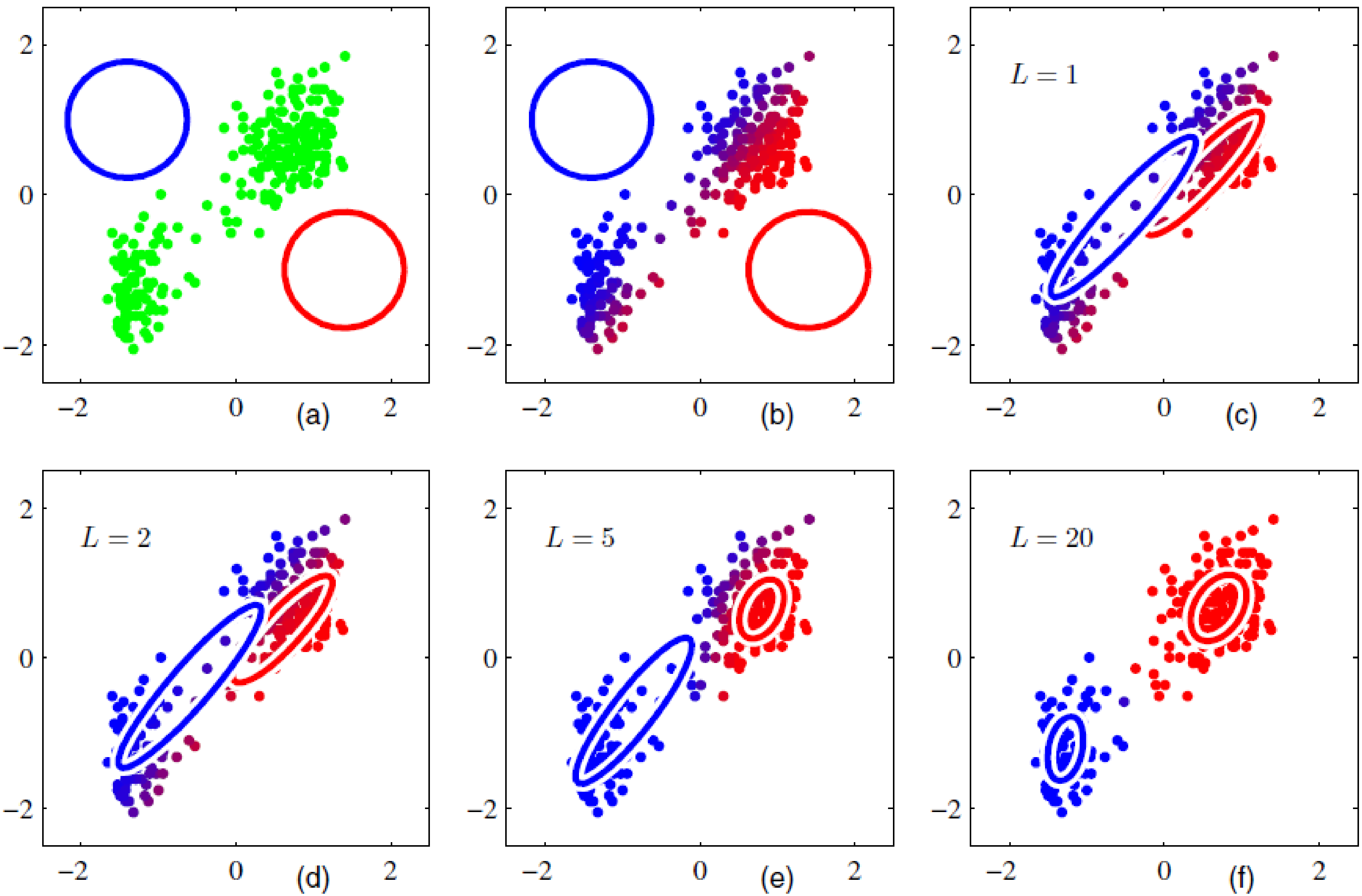
$$\mathbf{u}_j = \frac{1}{N_j}\sum_{n=1}^{N}\gamma(z_{nj})\mathbf{x}_n$$

$$\mathbf{\Sigma}_j = \frac{1}{N_j}\sum_{n=1}^{N}\gamma(z_{nj})(\mathbf{x}_n - \mathbf{u}_j)(\mathbf{x}_n - \mathbf{u}_j)^T$$

$$\pi_j = \frac{N_j}{N}$$

$$N_j = \sum_{n=1}^{N}\gamma(z_{nj})$$

EM for GMMs



Aut

Example

- (a) – initialization
- (b) – 1st E-step
- (c) – 1st M-step
- (d) to (f) 2nd, 5th and 20th EM step