More EM: Gaussian Mixture Models

CISC 5800
Professor Daniel Leeds

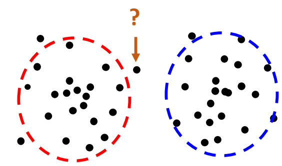
Clustering (generally unsupervised learning)

Group data points based on features

- E.g., k-means, hierarchical

Hard cluster:

- Each data point belongs to one cluster

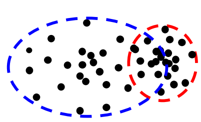


Soft/fuzzy cluster:

- Probability each data belongs to each cluster

Cluster challenges

- What if clusters overlap?
- What if clusters have different shapes?



Gaussian mixture models

The entire data set seen as a mixture of K clusters: $C_1, \ldots C_K$

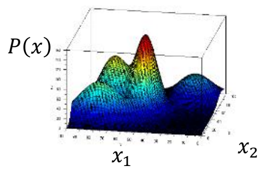
Prior probabilities: $p(C_k) = \pi_k \qquad \sum_k \pi_k = 1$

Gaussian likelihood for belonging in each cluster:

$$p(x^i|C_k) \sim N(x^i|\mu_k, \Sigma_k)$$

p(x) defined by mix of Gaussians

$P(\boldsymbol{x}) = \sum_k \pi_k N(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$



Goal: find π , $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$

Objective Function

$$\prod_i \sum_k \boldsymbol{\pi}_k N(\boldsymbol{x}^i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

$$\Sigma_k = \begin{bmatrix} \sigma_1^{2,k} & \sigma_{21}^k \\ \sigma_{12}^k & \sigma_2^{2,k} \end{bmatrix}$$

Expectation Maximization revisited

- E-step: compute expected cluster memberships for all data points
- M-step: compute likelihood parameters for each cluster

E-step

- Compute $P(C_k|\boldsymbol{x})$ given $P(\boldsymbol{x}|C_k)$ and π_k

$$P(C_k|\boldsymbol{x}) = \frac{\pi_k N(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_j \pi_j N(\boldsymbol{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}$$

M-step Define: $\gamma_{ik} = P(C_k|\boldsymbol{x})$

- Compute π_k $\quad N'_k = \sum_i \gamma_{ik}$, $\quad \pi_k = \frac{N'_k}{\sum_j N'_j}$
- Compute $\boldsymbol{\mu_k}$ $\quad \boldsymbol{\mu_k} = \frac{\sum_i \gamma_{ik} x^i}{N'_k}$
- Compute Σ_k $\quad \sigma_j^{2,k} = \frac{\sum_i \gamma_{ik}\left(x_j^i - \mu_{j,k}\right)^2}{N'_k}$