Gaussian Mixture Model (GMM) - standard construction

μ_k: mean
σ_k: covariance

A linear superposition of K-Gaussians

$$p(\mathrm{x}_i) = \sum_{k=1}^{K} \underbrace{\pi_k}_{p(k)} \underbrace{\mathcal{N}(\mathrm{x}_i|\mu_k, \sigma_k)}_{p(\mathrm{x}_i|k)}, \quad i = 1, \dots, N$$

is called a **Gaussian mixture (GM)**. The mixture coefficient π_k satisfies

$$\sum_{k=1}^{K} \pi_k = 1, \qquad 0 \leq \pi_k \leq 1$$

Interpretation: The density $p(\mathrm{x}|k) = \mathcal{N}(\mathrm{x}|\mu_k, \sigma_k)$ is the probability of x, given that component k was chosen. The probability of choosing component k is given by the prior probability $p(k)$.

GMM - standard construction (cont.)

For example, consider the following GMM:

$$p(x) = \underbrace{0.3}_{\pi_1} \mathcal{N}\left(x \mid \underbrace{\begin{pmatrix} 4 \\ 4.5 \end{pmatrix}}_{\mu_1}, \underbrace{\begin{pmatrix} 1.2 & 0.6 \\ 0.6 & 0.5 \end{pmatrix}}_{\Sigma_1}\right) + \underbrace{0.5}_{\pi_2} \mathcal{N}\left(x \mid \underbrace{\begin{pmatrix} 8 \\ 1 \end{pmatrix}}_{\mu_2}, \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}_{\Sigma_2}\right) + \underbrace{0.2}_{\pi_3} \mathcal{N}\left(x \mid \underbrace{\begin{pmatrix} 9 \\ 8 \end{pmatrix}}_{\mu_3}, \underbrace{\begin{pmatrix} 0.6 & 0.5 \\ 0.5 & 1.5 \end{pmatrix}}_{\Sigma_3}\right)$$

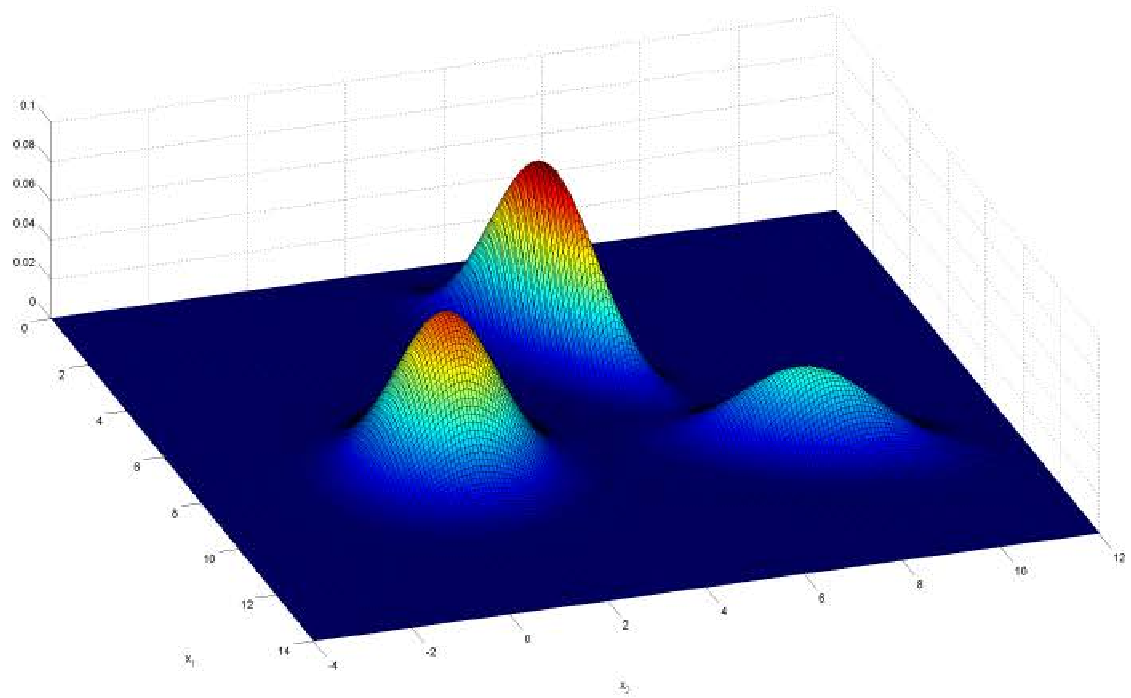

Figure: Probability density function.

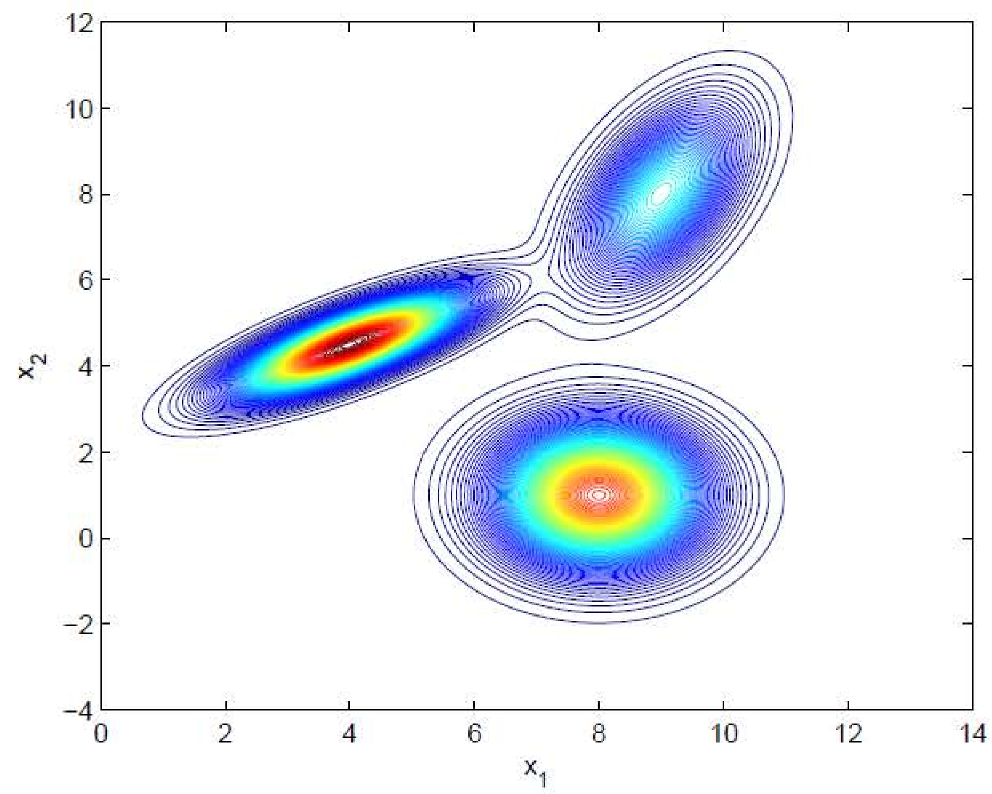

Figure: Contour plot.

GMM - standard construction (cont.)

The form of the GM distribution is governed by the parameters $\boldsymbol{\pi}, \boldsymbol{\mu}$ and $\boldsymbol{\sigma}$. One way to get them is by **maximum likelihood**.

Given N observations $\{\mathrm{x}_n\}_{n=1}^{N}$, the log-likelihood function is

$$\ln p\left(X; \pi_{1:K}, \mu_{1:K}, \sigma_{1:K}\right) = \sum_{n=1}^{N} \ln\left(\sum_{k=1}^{K} \pi_k \mathcal{N}\left(\mathrm{x}_n | \mu_k, \sigma_k\right)\right)$$

There is **no closed-form solution** available (due to the sum inside the logarithm).

This problem can be separated into two simple problems using the *expectation-maximization (EM)* algorithm.

GMM - standard construction (cont.)

Conditions to be satisfied at a maximum of the likelihood function

$$\frac{\mathrm{d}}{\mathrm{d}\mu_k}\left[\ln p(\mathrm{x}|\pi,\mu,\sigma)\right]=0 \quad \rightarrow \quad 0=-\sum_{n=1}^{N}\underbrace{\frac{\pi_k\mathcal{N}(\mathrm{x}_n|\mu_k,\sigma_k)}{\sum_j \pi_j\mathcal{N}(\mathrm{x}_n|\mu_j,\sigma_j)}}_{\gamma(z_{nk})}\sigma_k(\mathrm{x}_n-\mu_k)$$

which gives $\rightarrow \quad \mu_k=\frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})\mathrm{x}_n$

$$\frac{\mathrm{d}}{\mathrm{d}\sigma_k}\left[\ln p(\mathrm{x}|\pi,\mu,\sigma)\right]=0 \quad \rightarrow \quad \sigma_k=\frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})\left(\mathrm{x}_n-\mu_k\right)\left(\mathrm{x}_n-\mu_k\right)^T$$

Maximize $\ln p(\mathrm{x}|\pi,\mu,\sigma)$ with respect to π_k (using Lagrange multipliers) gives

$$\pi_k=\frac{N_k}{N}, \quad \text{where} \quad N_k=\sum_{n=1}^{N}\gamma(z_{nk})$$

For more details of EM and GMM see: C. Bishop, Pattern Recognition and Machine Learning, Springer, 2007.

GMM - standard construction (cont.)

Algorithm 1 EM for Gaussian mixtures

Initialize $\mu_k^1, \sigma_k^1, \pi_k^1$ and set $i = 1$.
while not converged **do**
  Compute $\gamma(z_{nk})$. ▷ Expectation step
  Compute $\mu_k^{i+1}; \pi_k^{i+1}; N_k; \sigma_k^{i+1}$. ▷ Maximization step
  $i \leftarrow i + 1$.
end while

$$\gamma(z_{nk}) = \frac{\pi_k^i \mathcal{N}(\mathrm{x}_n | \mu_k^i, \sigma_k^i)}{\sum_{j=1}^{K} \pi_j^i \mathcal{N}(\mathrm{x}_n | \mu_j^i, \sigma_j^i)}, n = 1, ..., N; k = 1, ..., K$$

$$\mu_k^{i+1} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \mathrm{x}_n,$$

$$\pi_k^{i+1} = \frac{N_k}{N}, \quad N_k = \sum_{n=1}^{N} \gamma(z_{nk}),$$

$$\sigma_k^{i+1} = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \left(\mathrm{x}_n - \mu_k^{i+1}\right) \left(\mathrm{x}_n - \mu_k^{i+1}\right)^T.$$