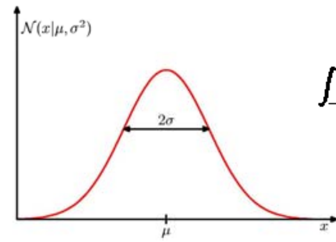
Mixture Models and EM

Chen Yu
Indiana University

The Gaussian Distribution

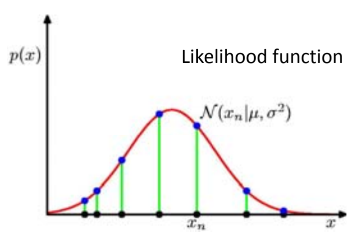
$$\mathcal{N}\left(x|\mu,\sigma^2\right) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$



$$\mathcal{N}\left(x|\mu,\sigma^2\right) > 0$$

$$\int_{-\infty}^{\infty} \mathcal{N}\left(x|\mu,\sigma^2\right) \, \mathrm{d}x = 1$$

Gaussian Parameter Estimation



$$p(\mathbf{x}|\mu,\sigma^2) = \prod_{n=1}^{N} \mathcal{N}\left(x_n|\mu,\sigma^2\right)$$

Maximize the likelihood function

- The likelihood function consists of two parameters.
- One common criterion for determining the parameters in a probability distribution using an observed data set is to find the parameter values that maximize the likelihood function.
- It is more convenient to maximize the log of the likelihood function because the logarithm is a monotonically increasing function of its argument.

Maximum (Log) Likelihood

$$p(X|\mu,\sigma^2) = \prod_{n=1}^{N} \frac{1}{(2\pi\sigma^2)^{1/2}} exp\left\{-\frac{1}{2\sigma^2}(x_n-\mu)^2\right\}$$

$$\ln p(X|\mu,\sigma^2) = -\frac{N}{2} ln\sigma^2 - \frac{N}{2}\ln(2\pi) - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x_n-\mu)^2$$

$$\mu_{ML} = \frac{1}{N}\sum_{n=1}^{N} x_n \qquad \sigma^2_{ML} = \frac{1}{N}\sum_{n=1}^{N}(x_n-\mu_{ML})^2$$

Mixtures of Gaussians

- We cannot use one single Gaussian distribution to model real data sets.
- A linear combination of Gaussians can give rise to very complex densities.

•By using a sufficient number of Gaussians, and by adjusting their means and covariances as well as the coefficients in the linear combination, almost any continuous density can be approximated.

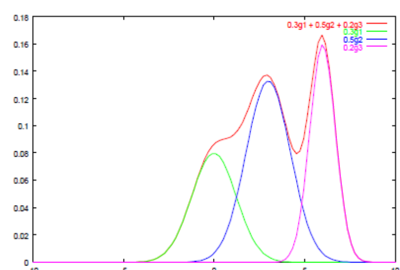
K Gaussian mixture

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

- Each Gaussian density is called a component of the mixture and has its own mean and covariance.
- The parameter π_k are called mixing coefficients:

$$\sum_{k=1}^{K} \pi_k = 1$$

Gaussian Mixture

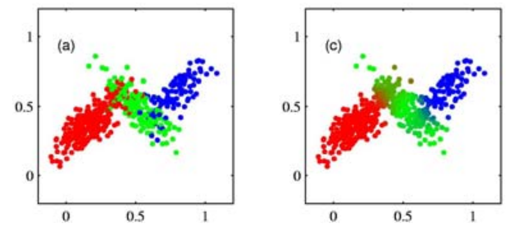


Posterior Probabilities

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

$$\textit{while } p(k) = \pi_k \textit{ and } p(x|k) = \mathcal{N}(x|\mu_k, \Sigma_k)$$

$$p(k|x) = \frac{p(k)p(x|k)}{p(x)} = \frac{p(k)p(x|k)}{\sum_{l=1}^{K} p(l)p(x|l)} = \frac{\pi_k \mathcal{N}(x|\mu_k, \Sigma_k)}{\sum_{l=1}^{K} \pi_l \mathcal{N}(x|\mu_l, \Sigma_l)}$$



The likelihood function of K Gaussian mixture

- The form of the Gaussian mixture distribution is governed by the parameters (π_k, μ_k, Σ_k)

- The likelihood function is given by:

$$\ln P(X|\pi, \mu, \Sigma) = \sum_{n=1}^{N} \ln\{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_n|\mu_k, \Sigma_k)\}$$

- The maximum likelihood solution no longer has a closed-form solution due to the presence of the summation over k inside the logarithm.

Maximizing the likelihood function

$$\frac{d \ln P(X|\pi, \mu, \Sigma)}{d\mu_k} = 0$$

$$\sum_{n=1}^{N} \frac{d \ln \sum_{j=1}^{K} \pi_j \mathcal{N}(x_n|\mu_j, \Sigma_j)\}}{d\mu_k} = 0$$

$$\sum_{n=1}^{N} \frac{1}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_n|\mu_j, \Sigma_j)\}} \frac{d \sum_{j=1}^{K} \pi_j \mathcal{N}(x_n|\mu_j, \Sigma_j)\}}{d\mu_k} = 0$$

$$\frac{d\sum_{j=1}^{K}\pi_j \mathcal{N}(x_n|\mu_j,\Sigma_j)\}}{d\mu_k} = \frac{d\,\pi_k\mathcal{N}(x_n|\mu_k,\Sigma_k)\}}{d\mu_k}$$

$$\frac{d\,\mathcal{N}(x_n|\mu_k,\Sigma_k)\}}{d\mu_k} = \frac{d\,\frac{1}{2\pi|\Sigma_k|^{1/2}}e^{(-\frac{1}{2}(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k))}}{d\mu_k}$$

$$= \frac{1}{2\pi|\Sigma_k|^{\frac{1}{2}}}e^{(-\frac{1}{2}(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k))}(x_n-\mu_k)\Sigma_k$$

$$= \mathcal{N}(x_n|\mu_k,\Sigma_k)(x_n-\mu_k)\Sigma_k$$

$$\sum_{n=1}^{N}\frac{1}{\sum_{j=1}^{K}\pi_j\mathcal{N}(x_n|\mu_j,\Sigma_j)\}}\,\frac{d\sum_{j=1}^{K}\pi_j\mathcal{N}(x_n|\mu_j,\Sigma_j)\}}{d\mu_k}=0$$

$$= \pi_k\mathcal{N}(x_n|\mu_k,\Sigma_k)(x_n-\mu_k)\Sigma_k$$

$$\sum_{n=1}^{N}\frac{\pi_k\mathcal{N}(x_n|\mu_k,\Sigma_k)\}}{\sum_{j=1}^{K}\pi_j\mathcal{N}(x_n|\mu_j,\Sigma_j)\}}(x_n-\mu_k)=0$$

$$\sum_{n=1}^{N}p(k|x_n)(x_n-\mu_k)=0$$

$$\mu_k = \frac{\sum_{n=1}^{N}p(k|x_n)x_n}{\sum_{n=1}^{N}p(k|x_n)}$$

$$\Sigma_k = \frac{\sum_{n=1}^{N}p(k|x_n)(x_n-\mu_k)(x_n-\mu_k)^T}{\sum_{n=1}^{N}p(k|x_n)}$$

$$\ln P(X|\pi,\mu,\Sigma) = \sum_{n=1}^{N}\ln\{\sum_{k=1}^{K}\pi_k\mathcal{N}(x_n|\mu_k,\Sigma_k)\} + \lambda(\sum_{k=1}^{K}\pi_k - 1)$$

For k cluster: $\sum_{n=1}^{N}\frac{\mathcal{N}(x_n|\mu_k,\Sigma_k)\}}{\sum_{j=1}^{K}\pi_j\mathcal{N}(x_n|\mu_j,\Sigma_j)\}} + \lambda = 0$

$$\sum_{n=1}^{N}\frac{\pi_k\mathcal{N}(x_n|\mu_k,\Sigma_k)\}}{\sum_{j=1}^{K}\pi_j\mathcal{N}(x_n|\mu_j,\Sigma_j)\}} + \lambda\pi_k = 0$$

$$N + \lambda = 0$$

$$\pi_k = \frac{\sum_{n=1}^{N}\frac{\pi_k\mathcal{N}(x_n|\mu_k,\Sigma_k)\}}{\sum_{j=1}^{K}\pi_j\mathcal{N}(x_n|\mu_j,\Sigma_j)\}}}{N} = \frac{\sum_{n=1}^{N}p(k|x_n)}{N}$$

EM Algorithm

- Step 0: Initialize parameters
- Step 1: E-step

$$p(k|x) = \frac{p(k)p(x|k)}{p(x)} = \frac{p(k)p(x|k)}{\sum_{l=1}^{K} p(l)p(x|l)} = \frac{\pi_k \mathcal{N}(x|\mu_k, \Sigma_k)}{\sum_{l=1}^{K} \pi_l \mathcal{N}(x|\mu_l, \Sigma_l)}$$

- Step 2: M-step

$$\mu_k = \frac{\sum_{n=1}^{N} p(k|x_n) x_n}{\sum_{n=1}^{N} p(k|x_n)} \quad \Sigma_k = \frac{\sum_{n=1}^{N} p(k|x_n)(x_n - \mu_k)(x_n - \mu_k)^T}{\sum_{n=1}^{N} p(k|x_n)}$$

$$\pi_k = \frac{\sum_{n=1}^{N} p(k|x_n)}{N}$$

- Step 3: check for converge?