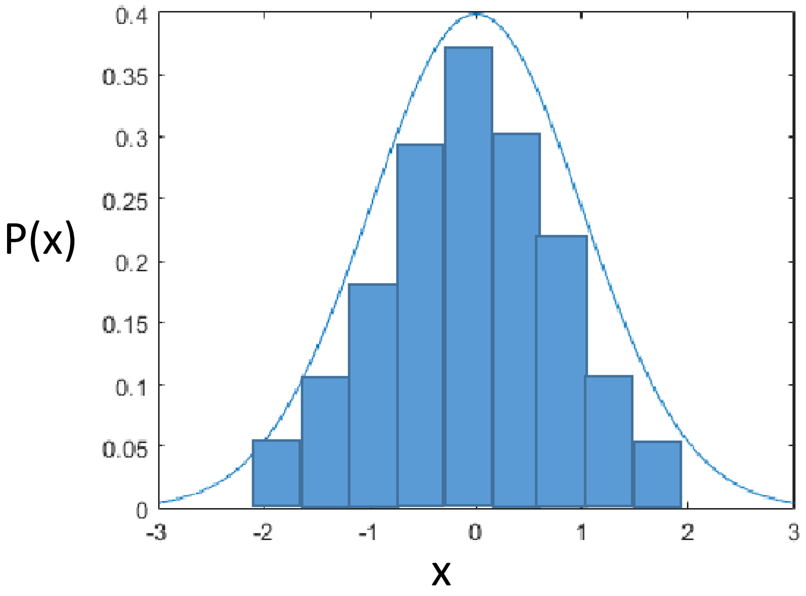
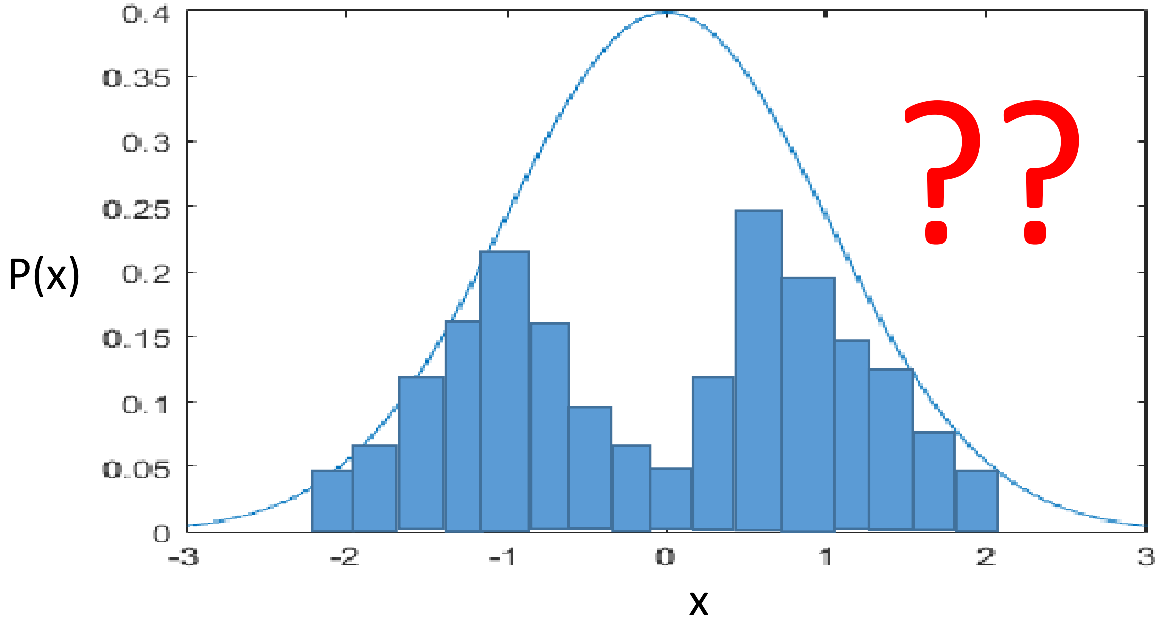
MIXTURE MODELS

MIXTURE MODELS

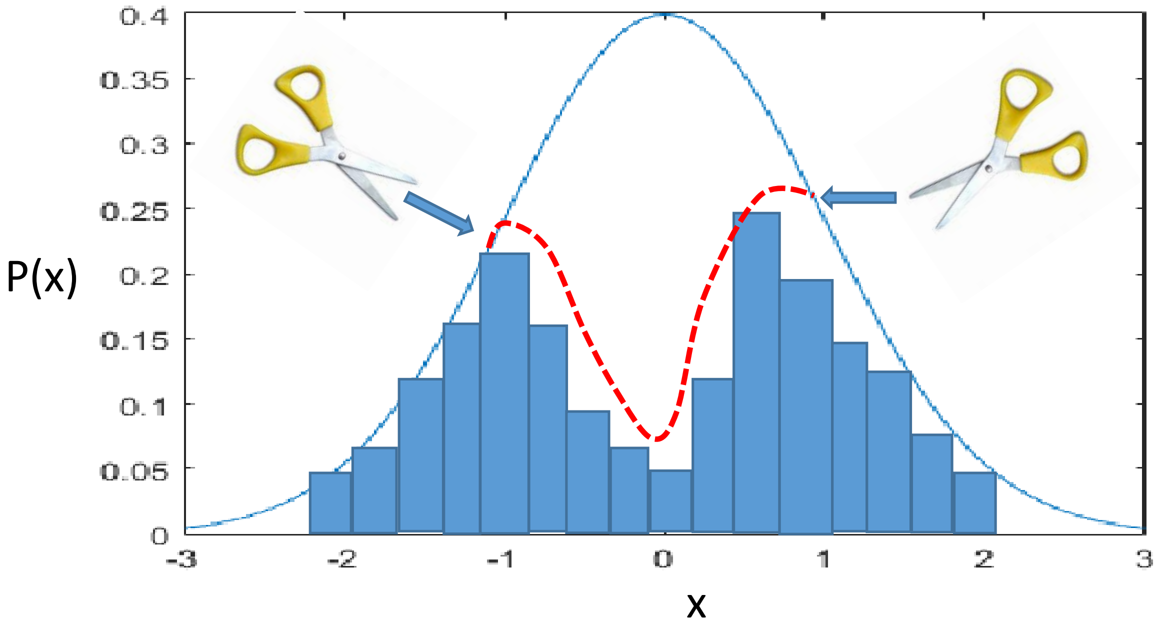
- Standard densities in many cases fail to capture the underlying probability distribution.

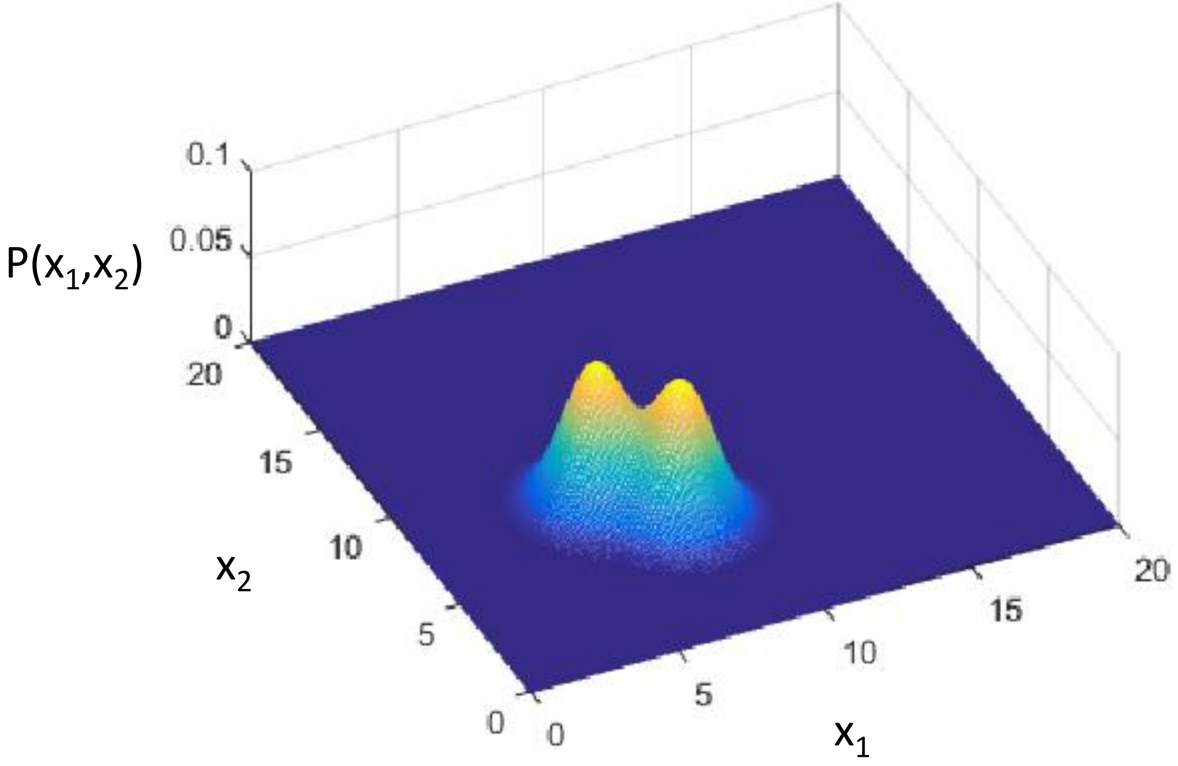




MIXTURE MODELS

- Modeling the density as a mixture of densities can be quite useful if individual densities fail to capture the data's behavior.





MIXTURE DENSITY MODEL

If each $f_i(x)$ *is a density function, then the density model is given as*

$$f(x) = \sum_{i=1}^{n} \lambda_i f_i(x)$$

where

$$\lambda_i \geq 0 \; and \; \sum_{i=1}^{n} \lambda_i = 1$$

The Gaussian Distribution

- Univariate Gaussian distribution

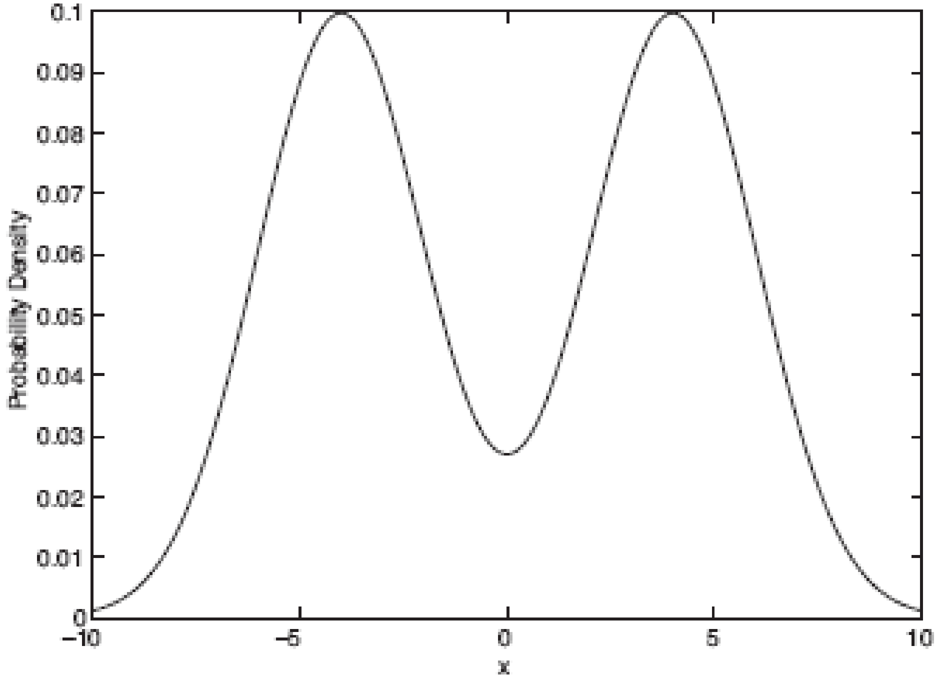
$$p(x|\mu,\sigma) = \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

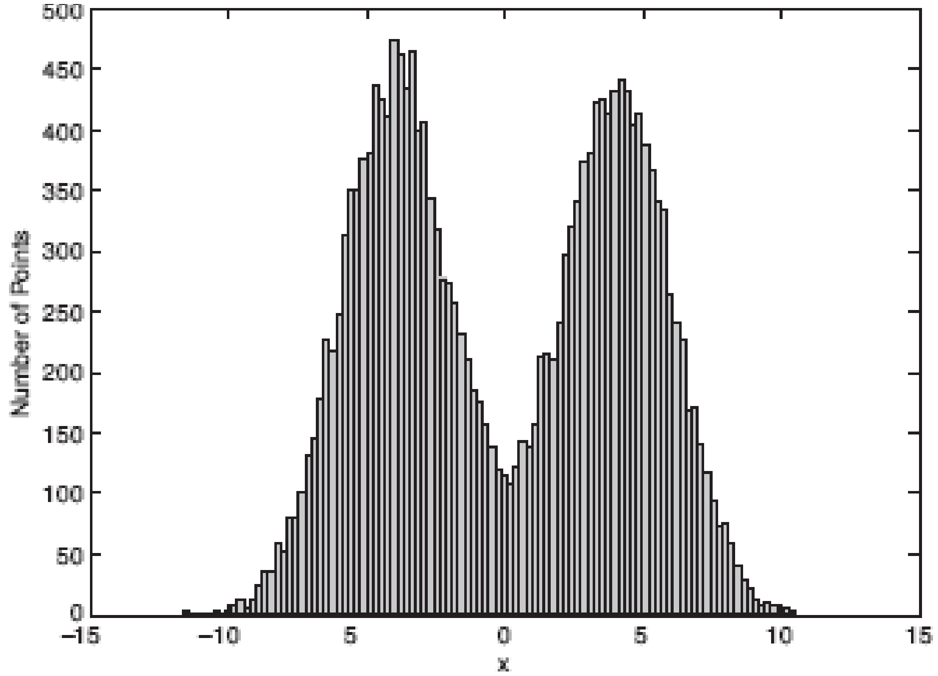
- Multivariate Gaussian distribution

$$p(x|\mu,\Sigma) = \frac{1}{((2\pi)^k|\Sigma|)^{\frac{1}{2}}} e^{\{\frac{-1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\}}$$

GMMs

- Example of a 1D GMM with 2 components is:
- $P(x) = w_1 * \frac{1}{(2\pi{\sigma_1}^2)^{\frac{1}{2}}} e^{\frac{-(x-\mu_1)^2}{2\sigma^2}} + w_2 * \frac{1}{(2\pi{\sigma_2}^2)^{\frac{1}{2}}} e^{\frac{-(x-\mu_2)^2}{2{\sigma_2}^2}}$
- Assuming $w_1 = w_2 = 0.5$ and μ_1=-4, μ_2=4 and $\sigma_1 = \sigma_2 = 2$
- then we have





Estimating GMM Parameters – MLE ?

- Lets try MLE and see how it goes …
- Let D={x_1, x_2, x_3, …, x_n} be a sample of iid data from this density

$$L(\theta|D) = \prod_{i=1}^{n}(\sum_{k=1}^{K} \lambda_k f_k(x_i))$$

$$LL(\theta|D) = \ln\left(\prod_{i=1}^{n}\left(\sum_{k=1}^{K} \lambda_k f_k(x_i)\right)\right)$$

$$= ln\sum_{k=1}^{K} \lambda_k f_k(x_1) + ln\sum_{k=1}^{K} \lambda_k f_k(x_2) + \cdots + ln\sum_{k=1}^{K} \lambda_k f_k(x_n)$$

$$= \sum_{i=1}^{n} ln\sum_{k=1}^{K} \lambda_k f_k(x_i)$$

Estimating GMM Parameters – MLE ?

- For simplicity, lets first examine the case of a mixture of two single dimensional Gaussians.
- Each individual Gaussian j in the mixture is:
- $\phi(x|\theta_j) = \frac{1}{\sigma_{j\sqrt{2\pi}}} e^{\frac{-(x-\mu_j)^2}{2\sigma_j^2}}$ $where\ \theta_j = (\mu_j, \sigma_j)$
- For our 2 components, if we substitute into the mixture formula, we have: $f(x|\theta) = \lambda_1 \phi(x|\theta_1) + \lambda_2 \phi(x|\theta_2)$
- Recall we had $\mathrm{LL}(\theta) = \sum_{i=1}^{n} ln \sum_{k=1}^{K} \lambda_k f_k(x_i)$
- This means $\mathrm{LL}(\theta) = \sum_{i=1}^{n} ln(\lambda_1 \phi(x_i|\theta_1) + \lambda_2 \phi(x_i|\theta_2))$

Estimating GMM Parameters – MLE ?

- $\text{LL}(\theta) = \sum_{i=1}^{n} ln(\lambda_1 \phi(x_i|\theta_1) + \lambda_2 \phi(x_i|\theta_2))$
- Maximizing the Log-likelihood function wrt parameters of mixture component #1

$$\frac{\partial \text{LL}(\theta)}{\partial \mu_1} = \frac{\partial}{\partial \mu_1}(\sum_{i=1}^{n} ln(\lambda_1 \phi(x_i|\theta_1) + \lambda_2 \phi(x_i|\theta_2)))$$

$$= \sum_{i=1}^{n} \frac{\partial}{\partial \mu_1} l\, n(\lambda_1 \phi(x_i|\theta_1) + \lambda_2 \phi(x_i|\theta_2)) = 0$$

- Similarly,

$$\frac{\partial \text{LL}(\theta)}{\partial \sigma_1} = \sum_{i=1}^{n} \frac{\partial}{\partial \sigma_1} l\, n(\lambda_1 \phi(x_i|\theta_1) + \lambda_2 \phi(x_i|\theta_2)) = 0$$

- Do the same for parameters of mixture component #2

Estimating GMM Parameters – MLE ?

- In preparation for maximization of likelihood function, lets check out the derivative of a Gaussian wrt (μ, θ)
- First recall that if we have y=$e^{f(x)}$, then dy/dx= $f'(x)\, e^{f(x)}$
- Recall; $\phi\left(x\middle|\theta_j\right) = \frac{1}{\sigma_{j\sqrt{2\pi}}} e^{\frac{-(x-\mu_j)^2}{2\sigma_j^2}}$
- $\frac{\partial}{\partial\mu_1}\phi(x|\theta_1) = \frac{\partial}{\partial\mu_1}\left(\frac{1}{\sigma_{1\sqrt{2\pi}}} e^{\frac{-(x-\mu_1)^2}{2\sigma_1^2}}\right) = -1 * 2 * \frac{-(x-\mu_1)}{2\sigma_1^2} \frac{1}{\sigma_{1\sqrt{2\pi}}} e^{\frac{-(x-\mu_1)^2}{2\sigma_1^2}}$
- $\Rightarrow \frac{\partial}{\partial\mu_1}\phi(x|\theta_1) = \phi(x|\theta_1).\frac{(x-\mu_1)}{\sigma_1^2}$

Estimating GMM Parameters – MLE ?

- $\frac{\partial}{\partial \sigma_1} \phi(x|\theta_1) = \frac{\partial}{\partial \sigma_1}\left(\frac{1}{\sigma_{1\sqrt{2\pi}}} e^{\frac{-(x-\mu_1)^2}{2\sigma_1^2}}\right) =$
- Applying the product rule with u= $\frac{1}{\sigma_{1\sqrt{2\pi}}}$ and v=$e^{\frac{-(x-\mu_1)^2}{2\sigma_1^2}}$:
- $\frac{\partial}{\partial \sigma_1} \phi(x|\theta_1) = \frac{-\sigma_1^{-2}}{\sqrt{2\pi}} . e^{\frac{-(x-\mu_1)^2}{2\sigma_1^2}} + \frac{1}{\sigma_{1\sqrt{2\pi}}} . \frac{-(x-\mu_1)^2 * -2\sigma_1^{-3}}{2} e^{\frac{-(x-\mu_1)^2}{2\sigma_1^2}}$
- $= \frac{1}{\sigma_{1\sqrt{2\pi}}} e^{\frac{-(x-\mu_1)^2}{2\sigma_1^2}} (-\frac{1}{\sigma_1} + \frac{(x-\mu_1)^2}{\sigma_1^3})$
- $\Rightarrow \frac{\partial}{\partial \sigma_1} \phi(x|\theta_1) = \phi(x|\theta_1). (-\frac{1}{\sigma_1} + \frac{(x-\mu_1)^2}{\sigma_1^{-3}})$

Estimating GMM Parameters – MLE ?

- Please don't forget that we had only branched off from our main task of evaluating:

$$\frac{\partial \mathrm{LL}(\theta)}{\partial \mu_1} = \sum_{i=1}^{n} \frac{\partial}{\partial \mu_1} l\,n\left(\lambda_1 \phi(x_i|\theta_1) + \lambda_2 \phi(x_i|\theta_2)\right) = 0 \text{ and}$$

$$\frac{\partial \mathrm{LL}(\theta)}{\partial \sigma_1} = \sum_{i=1}^{n} \frac{\partial}{\partial \sigma_1} l\,n\left(\lambda_1 \phi(x_i|\theta_1) + \lambda_2 \phi(x_i|\theta_2)\right) = 0$$

Since we now have the derivatives $\frac{\partial}{\partial \mu_1}\phi(x|\theta_1)$ and $\frac{\partial}{\partial \sigma_1}\phi(x|\theta_1)$, we are well placed to move on…

$$\sum_{i=1}^{n} \frac{\partial}{\partial \mu_1} l\,n\left(\lambda_1 \phi(x_i|\theta_1) + \lambda_2 \phi(x_i|\theta_2)\right) = \sum_{i=1}^{n} \frac{\lambda_1 \phi(x_i|\theta_1).\frac{(xi-\mu_1)}{\sigma_1^2}}{\lambda_1 \phi(x_i|\theta_1) + \lambda_2 \phi(x_i|\theta_2)}$$

Estimating GMM Parameters – MLE ?

- Similarly,
- $\sum_{i=1}^{n} \frac{\partial}{\partial \sigma_1} l\, n\left(\lambda_1 \phi(x_i|\theta_1) + \lambda_2 \phi(x_i|\theta_2)\right) = \sum_{i=1}^{n} \frac{\lambda_1 \phi(x_i|\theta_1).(\frac{(x-\mu_1)^2}{\sigma_1^3} - \frac{1}{\sigma_1})}{\lambda_1 \phi(x_i|\theta_1) + \lambda_2 \phi(x_i|\theta_2)}$
- If we define $\gamma_{i1} = \frac{\lambda_1 \phi(x_i|\theta_1)}{\lambda_1 \phi(x_i|\theta_1) + \lambda_2 \phi(x_i|\theta_2)}$
- We can write: $\frac{\partial \mathrm{LL}(\theta)}{\partial \sigma_1} = \sum_{i=1}^{n} \gamma_{i1}. (\frac{(x-\mu_1)^2}{\sigma_1^3} - \frac{1}{\sigma_1})\ and$
- $\frac{\partial \mathrm{LL}(\theta)}{\partial \mu_1} = \sum_{i=1}^{n} \gamma_{i1}. \frac{(xi-\mu_1)}{\sigma_1^2}\ and$

Estimating GMM Parameters – MLE ?

- Now we have the derivatives, lets set them to zero …
- $\sum_{i=1}^{n} \gamma_{i1} . \left(\frac{(x-\mu_1)^2}{\sigma_1^3} - \frac{1}{\sigma_1} \right) = 0 \Rightarrow \sum_{i=1}^{n} \gamma_{i1} . \left(\frac{(x-\mu_1)^2 - \sigma_1^2}{\sigma_1^3} \right) = 0$
- $\frac{1}{\sigma_1^3} \sum_{i=1}^{n} \gamma_{i1} . \left((x-\mu_1)^2 - \sigma_1^2 \right) = 0 \Rightarrow \sum_{i=1}^{n} \gamma_{i1} . \left((x-\mu_1)^2 - \sigma_1^2 \right) = 0$

$$\sum_{i=1}^{n} \gamma_{i1} . \left((x-\mu_1)^2 - \sigma_1^2 \right) = 0 \Rightarrow \sum_{i=1}^{n} \gamma_{i1} . \left((x-\mu_1)^2 \right) = \sigma_1^2 \sum_{i=1}^{n} \gamma_{i1}$$

$$\sigma_1^2 = \frac{\sum_{i=1}^{n} \gamma_{i1} . \left((x-\mu_1)^2 \right)}{\sum_{i=1}^{n} \gamma_{i1}}$$

Estimating GMM Parameters – MLE ?

- Continue setting the derivatives to zero …
- $\sum_{i=1}^{n} \gamma_{i1} . \frac{(xi-\mu_1)}{\sigma_1^2} = 0 \Rightarrow \frac{1}{\sigma_1^2} \sum_{i=1}^{n} \gamma_{i1} (xi - \mu_1) = 0$
- $\sum_{i=1}^{n} \gamma_{i1} . xi = \sum_{i=1}^{n} \gamma_{i1} \mu_1 \Rightarrow \sum_{i=1}^{n} \gamma_{i1} . xi = \mu_1 \sum_{i=1}^{n} \gamma_{i1}$

$$\mu_1 = \frac{\sum_{i=1}^{n} \gamma_{i1} . x_i}{\sum_{i=1}^{n} \gamma_{i1}}$$

GMM Parameters – MLE ?

- Putting it all together …

- $\sigma_1^2 = \frac{\sum_{i=1}^{n} \gamma_{i1}.((xi - \mu_1)^2)}{\sum_{i=1}^{n} \gamma_{i1}}$
- $\mu_1 = \frac{\sum_{i=1}^{n} \gamma_{i1}.x_i}{\sum_{i=1}^{n} \gamma_{i1}}$
- Following the same procedure, we can easily get
- $\sigma_2^2 = \frac{\sum_{i=1}^{n} \gamma_{i2}.((xi - \mu_2)^2)}{\sum_{i=1}^{n} \gamma_{i2}}$
- $\mu_2 = \frac{\sum_{i=1}^{n} \gamma_{i2}.x_i}{\sum_{i=1}^{n} \gamma_{i2}}$

EM iterates for GMM

- $(\sigma_j^2)^{(k+1)} = \frac{\sum_{i=1}^{n} \gamma_{ij(k)} \cdot \left((xi - \mu_{j(k)})^2 \right)}{\sum_{i=1}^{n} \gamma_{ij(k)}}$
- $\mu_{j(k+1)} = \frac{\sum_{i=1}^{n} \gamma_{ij(k)} \cdot x_i}{\sum_{i=1}^{n} \gamma_{ij(k)}}$
- $\gamma_{ij(k+1)} = \frac{\lambda_{j(k)} \phi\left(x_i \middle| \theta_{j(k)}\right)}{\sum_{j=1}^{2} \lambda_{j(k)} \phi\left(x_i \middle| \theta_{j(k)}\right)}$
- $\lambda_j^{(k+1)} = \frac{1}{n} \sum_{i=1}^{n} \gamma_{ij(k)}$

Mixture Models

- The previous process can be generalized to any number of components and any identity of distributions.
- GMMS more common however
- How does one estimate the number of components?
  - Domain knowledge
  - Cross validation – run the process on different data subsets and evaluate performance