Tutorial on Expectation Maximization (Example)

Expectation Maximization (Intuition)
Expectation Maximization (Maths)

EM: the intuition

- Assume that we have two coins, C1 and C2
- Assume the bias of C1 is θ_1
  (i.e., probability of getting heads with C1)
- Assume the bias of C2 is θ_2
  (i.e., probability of getting heads with C2)

- We want to find θ_1, θ_2 by performing a number of trials
  (i.e., coin tosses)

EM: the intuition

First experiment

- We choose 5 times one of the coins.
- We toss the chosen coin 10 times

$$\theta_1 = \frac{number\ of\ heads\ using\ C1}{total\ number\ of\ flips\ using\ C1}$$

$$\theta_2 = \frac{number\ of\ heads\ using\ C2}{total\ number\ of\ flips\ using\ C2}$$

EM: the intuition

H T T T H H T H T H

H H H H T H H H H H

H T H H H H H T H H

H T H T T T H H T T

T H H H T H H H T H

Coin A	Coin B
	5 H, 5 T
9 H, 1 T	
8 H, 2 T	
	4 H, 6 T
7 H, 3 T	
24 H, 6 T	9 H, 11 T

$$\theta_1 = \frac{24}{24+6} = 0.8$$

$$\theta_2 = \frac{9}{9+11} = 0.45$$

EM: the intuition

Assume a more challenging problem

H T T T H H T H T H

H H H H T H H H H H

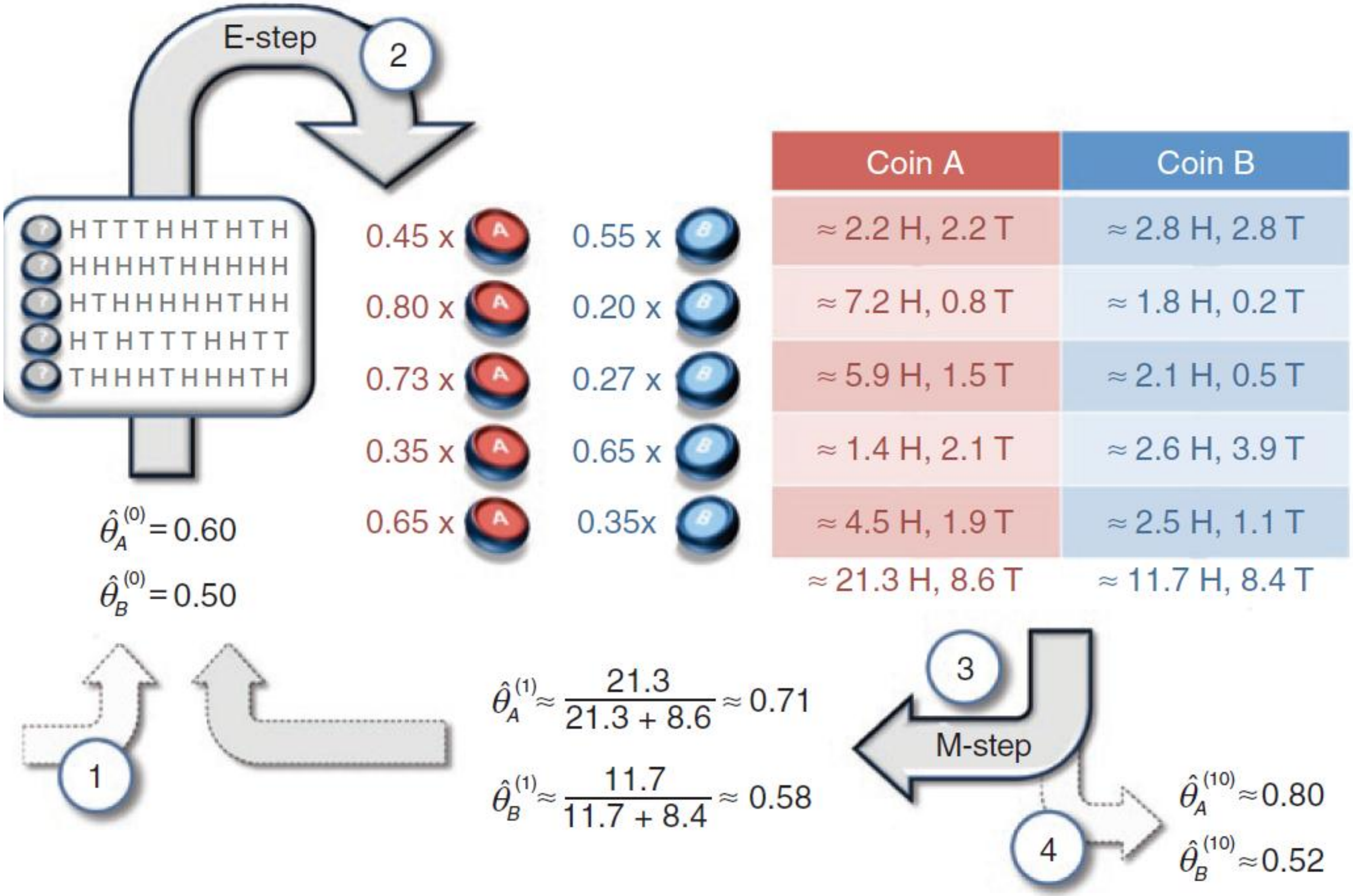
H T H H H H H T H H

H T H T T T H H T T

T H H H T H H H T H

•We do not know the identities of the coins used for each set of tosses (we treat them as hidden variables).

EM: the intuition



EM: the Maths (setting the joint)

$$\mathbf{z}_i = \begin{bmatrix} z_{i1} \\ z_{i2} \end{bmatrix} \in \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

$$p(X_1, X_2, \cdots, X_5, \mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_5 | \theta)$$
$$= p(\{x_1{}^1, \cdots, x_1{}^{10}\}, \cdots, \{x_5{}^1, \cdots, x_5{}^{10}\}, \mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_5 | \theta)$$

$$= p(\{x_1{}^1, \cdots, x_1{}^{10}\}, \cdots, \{x_5{}^1, \cdots, x_5{}^{10}\} | \mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_5, \theta)\, p(\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_5)$$
$$= \prod_{i=1}^{5} p(\{x_i{}^1, \cdots, x_i{}^{10}\} | \mathbf{z}_i, \theta) \prod_{i=1}^{5} p(\mathbf{z}_i)$$

$$p(\mathbf{z}_i) = \prod_{k=1}^{2} \pi_\kappa{}^{z_{ik}}$$ π_κ *is the probability of selecting coin* $k \in \{1,2\}$

$$p(\{x_i{}^1, \cdots, x_i{}^{10}\} | \mathbf{z}_i, \theta) = \prod_{j=1}^{10} p(x_i{}^j | \mathbf{z}_i, \theta)$$

EM: the Maths (setting the joint)

$x_i{}^j = 1$ *If* j *toss of* i *run is head*

$x_i{}^j = 0$ *If* j *toss of* i *run is head*

$$p\left(x_i{}^j \middle| \mathbf{z}_i, \theta\right) = \prod_{k=1}^{2} \left[\theta_k{}^{x_i{}^j} (1-\theta_k)^{1-x_i{}^j}\right]^{z_{ik}}$$

then

$$p\left(X_1, X_2, \cdots, X_5, \mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_5 \middle| \theta\right) = \prod_{i=1}^{5} \prod_{j=1}^{10} \prod_{k=1}^{2} \left[\theta_k{}^{x_i{}^j} (1-\theta_k)^{1-x_i{}^j}\right]^{z_{ik}} \prod_{i=1}^{5} \prod_{k=1}^{2} \pi_\kappa{}^{z_{ik}}$$

EM: the Maths (computing the expectation)

$$\ln p(X_1, X_2, \cdots, X_5, \boldsymbol{z}_{1,} \boldsymbol{z}_2, \cdots, \boldsymbol{z}_5 | \theta)$$
$$= \sum_{i=1}^{5} \sum_{j=1}^{10} \sum_{k=1}^{2} z_{ik} \ln \theta_k^{x_i{}^j} (1 - \theta_k)^{1 - x_i{}^j} + \sum_{i=1}^{5} \sum_{k=1}^{2} z_{ik} \ln \pi_\kappa$$

Taking the expectation of the above

$$E_{p(Z|X)}[\ln p(X_1, X_2, \cdots, X_5, \boldsymbol{z}_{1,} \boldsymbol{z}_2, \cdots, \boldsymbol{z}_5 | \theta)]$$
$$= \sum_{i=1}^{5} \sum_{j=1}^{10} \sum_{k=1}^{2} E_{p(Z|X)}[z_{ik}] \ln \theta_k^{x_i{}^j} (1 - \theta_k)^{1 - x_i{}^j}$$
$$+ \sum_{i=1}^{5} \sum_{k=1}^{2} E_{p(Z|X)}[z_{ik}] \ln \pi_\kappa$$

EM: Expectation Step

$$p(\boldsymbol{Z}|\boldsymbol{X},\theta) = p\left(\boldsymbol{z}_{1,}\boldsymbol{z}_2,\cdots,\boldsymbol{z}_5|X_1,X_2,\cdots,X_5,\theta\right)$$

$$E_{p(\boldsymbol{Z}|\boldsymbol{X})}[z_{ik}] = \sum_{\boldsymbol{z}_1}\cdots\sum_{\boldsymbol{z}_5} z_{ik}\, p(\boldsymbol{Z}|\boldsymbol{X},\theta) = \sum_{\boldsymbol{z}_i} z_{ik} p(\boldsymbol{z}_i|\{x_i{}^1,\cdots,x_i{}^{10}\})$$

$$p(\boldsymbol{z}_i|\{x_i{}^1,\cdots,x_i{}^{10}\}) = \frac{p(\{x_i{}^1,\cdots,x_i{}^{10}\}|\boldsymbol{z}_i,\theta)p(\boldsymbol{z}_i)}{p(\{x_i{}^1,\cdots,x_i{}^{10}\}|\theta)}$$

$$= \frac{\prod_{j=1}^{10}\prod_{k=1}^{2}\left[\theta_k{}^{x_i{}^j}(1-\theta_k)^{1-x_i{}^j}\right]^{z_{ik}}\pi_\kappa{}^{z_{ik}}}{\sum_{\boldsymbol{z}_i}\prod_{j=1}^{10}\prod_{k=1}^{2}\left[\theta_k{}^{x_i{}^j}(1-\theta_k)^{1-x_i{}^j}\right]^{z_{ik}}\pi_\kappa{}^{z_{ik}}}$$

EM: Expectation Step

$$E_{p(\boldsymbol{Z}|\boldsymbol{X})}[z_{ik}] = \sum_{\boldsymbol{z}_i} z_{ik} \frac{\prod_{j=1}^{10} \prod_{k=1}^{2} \left[\theta_k{}^{x_i{}^j} (1-\theta_k)^{1-x_i{}^j}\right]^{z_{ik}} \pi_\kappa{}^{z_{ik}}}{\sum_{\boldsymbol{z}_i} \prod_{j=1}^{10} \prod_{k=1}^{2} \left[\theta_k{}^{x_i{}^j} (1-\theta_k)^{1-x_i{}^j}\right]^{z_{ik}} \pi_\kappa{}^{z_{ik}}}$$

$$= \frac{\sum_{\boldsymbol{z}_i} z_{ik} \prod_{j=1}^{10} \prod_{k=1}^{2} \left[\theta_k{}^{x_i{}^j} (1-\theta_k)^{1-x_i{}^j}\right]^{z_{ik}} \pi_\kappa{}^{z_{ik}}}{\sum_{\boldsymbol{z}_i} \prod_{j=1}^{10} \prod_{k=1}^{2} \left[\theta_k{}^{x_i{}^j} (1-\theta_k)^{1-x_i{}^j}\right]^{z_{ik}} \pi_\kappa{}^{z_{ik}}}$$

$$= \frac{\pi_\kappa \prod_{j=1}^{10} \theta_k{}^{x_i{}^j} (1-\theta_k)^{1-x_i{}^j}}{\pi_1 \prod_{j=1}^{10} \theta_1{}^{x_i{}^j} (1-\theta_1)^{1-x_i{}^j} + \pi_2 \prod_{j=1}^{10} \theta_2{}^{x_i{}^j} (1-\theta_2)^{1-x_i{}^j}}$$

EM: Expectation Step

Expectation (E) Step (fix θ):

$$\pi_1 = \frac{1}{2} \qquad \pi_2 = \frac{1}{2}$$

$$E_{p(\boldsymbol{z}|\boldsymbol{x})}[z_{ik}] = \frac{\prod_{j=1}^{10} \theta_k^{\,x_i{}^j}(1-\theta_k)^{1-x_i{}^j}}{\prod_{j=1}^{10} \theta_1^{\,x_i{}^j}(1-\theta_1)^{1-x_i{}^j} + \prod_{j=1}^{10} \theta_2^{\,x_i{}^j}(1-\theta_2)^{1-x_i{}^j}}$$

EM: Maximization Step

Maximization Step (fix $E_{p(\mathbf{Z}|\mathbf{X})}[z_{ik}]$):

$$\max L(\theta) = E_{p(Z|X)}\left[\ln p\left(X_1, X_2, \cdots, X_5, \mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_5 | \theta\right)\right]$$

$$= \sum_{i=1}^{5}\sum_{j=1}^{10}\sum_{k=1}^{2} E_{p(Z|X)}[z_{ik}] \ln \theta_k^{x_i^j}(1-\theta_k)^{1-x_i^j} + \sum_{i=1}^{5}\sum_{k=1}^{2} E_{p(Z|X)}[z_{ik}] \ln \pi_\kappa$$

$$= \sum_{i=1}^{5}\sum_{j=1}^{10}\sum_{k=1}^{2} E_{p(Z|X)}[z_{ik}]\left(x_i^j \ln\theta_k + \left(1 - x_i^j\right)\ln(1-\theta_k)\right) + const$$

EM: Maximization Step

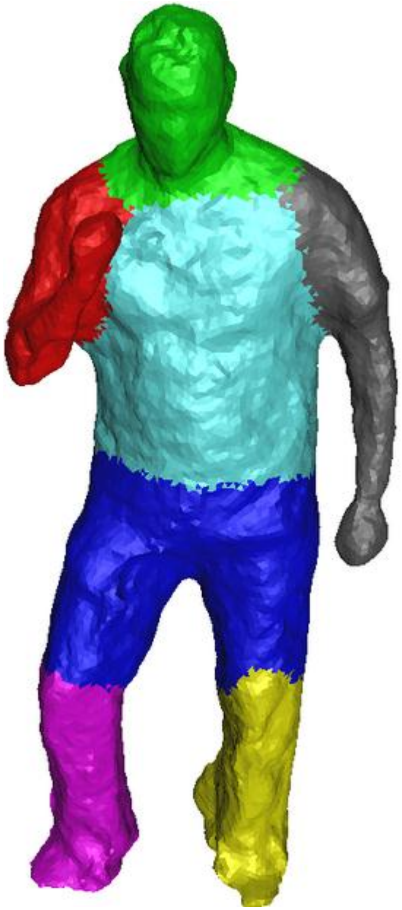
$$\frac{dL(\theta)}{d\theta_1} = \sum_{i=1}^{5}\sum_{j=1}^{10} E_{p(Z|X)}[z_{i1}]({x_i}^j \frac{1}{\theta_1} - \left(1 - {x_i}^j\right)\frac{1}{1-\theta_1}) = 0$$

$$\sum_{i=1}^{5}\sum_{j=1}^{10} E_{p(Z|X)}[z_{i1}]({x_i}^j(1-\theta_1) - \left(1 - {x_i}^j\right)\theta_1) = 0$$

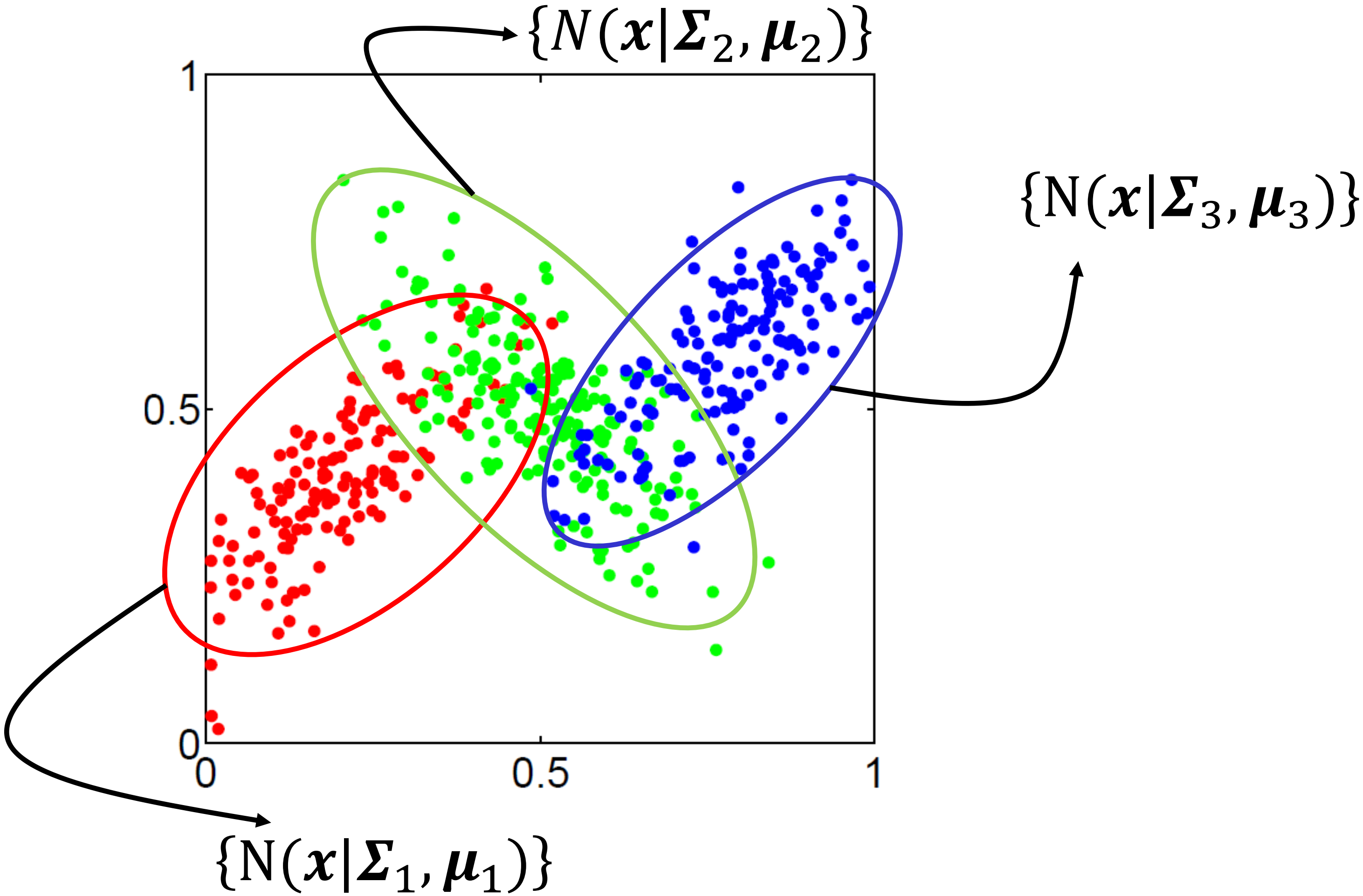
$$\theta_1 = \frac{\sum_{i=1}^{5}\sum_{j=1}^{10} E_{p(Z|X)}[z_{i1}]{x_i}^j}{\sum_{i=1}^{5} 10\, E_{p(Z|X)}[z_{i1}]} \qquad \theta_2 = \frac{\sum_{i=1}^{5}\sum_{j=1}^{10} E_{p(Z|X)}[z_{i2}]{x_i}^j}{\sum_{i=1}^{5} 10\, E_{p(Z|X)}[z_{i2}]}$$

Clustering

Clustering

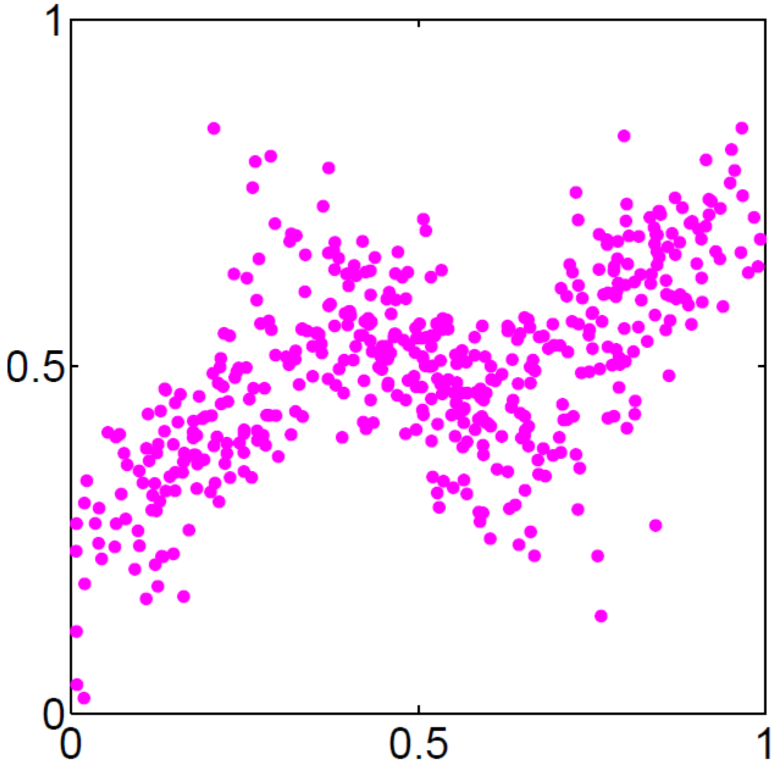


Gaussian Mixture Models



Gaussian Mixture Models

We are given a set of un-labelled data



We need to find

Parameters $\{\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \boldsymbol{\Sigma}_3, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}_3\}$

and also $p(k=1), p(k=2), p(k=3)$

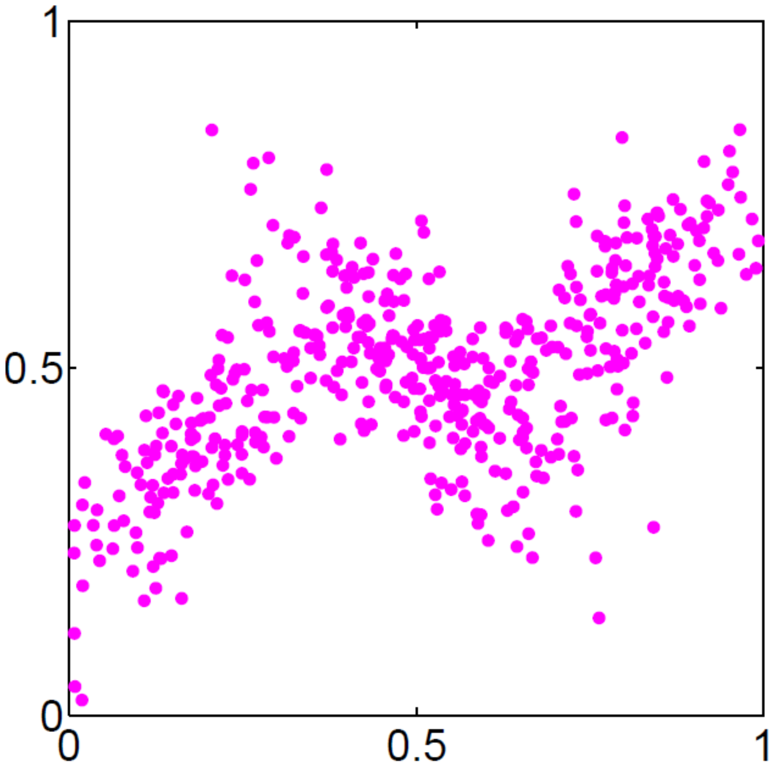
What are our hidden variables?

$$\boldsymbol{z}_n = \begin{bmatrix} z_{n1} \\ z_{n2} \\ z_{n3} \end{bmatrix} \in \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$$

$$\text{e.g., } \pi_1 = p(z_1 = 1) = p(k=1)$$

$$p(\boldsymbol{z}_n) = \prod_{k=1}^{3} \pi_\kappa{}^{z_{nk}}$$

Gaussian Mixture Models



$$p(\boldsymbol{x}_n | z_{nk} = 1, \theta) = N(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

$$p(\boldsymbol{x}_n | \boldsymbol{z}_n, \theta) = \prod_{k=1}^{3} N(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{z_{nk}}$$

The probability of a sample $\boldsymbol{x}_n$ is given by the sum rule:

$$p(\boldsymbol{x}_n | \theta) = \sum_{k=1}^{3} p(z_{nk} = 1)\, p(\boldsymbol{x}_n | z_{nk} = 1, \theta) = \sum_{k=1}^{3} \pi_k\, N(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \Sigma_k)$$

Gaussian Mixture Models

Assume all data samples are independent.

We, as always, formulate the joint likelihood.

$$p(\boldsymbol{X},\boldsymbol{Z}|\theta) = p\left(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_N,\boldsymbol{z}_{1,}\boldsymbol{z}_2,\cdots,\boldsymbol{z}_N|\theta\right)$$

$$= \prod_{n=1}^{N} p(\boldsymbol{x}_n|\boldsymbol{z}_n,\theta_x)\prod_{n=1}^{N} p(\boldsymbol{z}_n|\theta_z)$$

$$\theta_x = \{\boldsymbol{\Sigma}_1,\boldsymbol{\Sigma}_2,\boldsymbol{\Sigma}_3,\boldsymbol{\mu}_1,\boldsymbol{\mu}_2,\boldsymbol{\mu}_3\}$$

$$\theta_z = \{\pi_1,\pi_2,\pi_3\}$$

$$= \prod_{n=1}^{N}\prod_{k=1}^{3} N(\boldsymbol{x}_n|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)^{\,z_{nk}}\prod_{n=1}^{N}\prod_{k=1}^{3} {\pi_\kappa}^{z_{nk}}$$

Gaussian Mixture Models(Expectation Step)

$$\ln p(\boldsymbol{X}, \boldsymbol{Z}|\theta) = \sum_{n=1}^{N} \sum_{k=1}^{3} z_{nk} \{\ln N(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) + \ln \pi_\kappa\}$$

Applying operator $E_{p(\boldsymbol{Z}|\boldsymbol{X},\theta)}$

$$E_{p(\boldsymbol{Z}|\boldsymbol{X},\theta)}[\ln p(\boldsymbol{X}, \boldsymbol{Z}|\theta)] = \sum_{n=1}^{N} \sum_{k=1}^{3} E_{p(\boldsymbol{Z}|\boldsymbol{X},\theta)}[z_{nk}] \{\ln N(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) + \ln \pi_\kappa\}$$

We need to compute $E_{p(\boldsymbol{Z}|\boldsymbol{X},\theta)}[z_{nk}]$

Gaussian Mixture Models (Expectation Step)

$$E_{p(\boldsymbol{Z}|\boldsymbol{X},\theta)}[z_{nk}] = \sum_{z_1} \cdots \sum_{z_\mathrm{N}} z_{nk}\, p\left(\boldsymbol{Z} \middle| \boldsymbol{X}, \theta^{old}\right)$$

$$= \sum_{z_n} z_{nk} p\left(\boldsymbol{z}_n \middle| \boldsymbol{x}_n, \theta^{old}\right)$$

$$p\left(\boldsymbol{z}_n \middle| \boldsymbol{x}_n, \theta^{old}\right) = \frac{p(\boldsymbol{x}_n, \boldsymbol{z}_n | \theta^{old})}{p(\boldsymbol{x}_n | \theta^{old})} = \frac{p\left(\boldsymbol{x}_n \middle| \boldsymbol{z}_n, \theta^{old}\right) p(\boldsymbol{z}_n | \theta^{old})}{p(\boldsymbol{x}_n | \theta^{old})}$$

$$= \frac{\prod_{k=1}^{3} N(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{\,z_{nk}} {\pi_\kappa}^{z_{nk}}}{\sum_{z_{nk}} \prod_{k=1}^{3} N(\boldsymbol{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{\,z_{nk}} {\pi_\kappa}^{z_{nk}}}$$

Gaussian Mixture Models (Expectation Step)

$$E_{p(\mathbf{Z}|\mathbf{X},\theta)}[z_{nk}] = \sum_{\mathbf{z}_n} z_{nk} p\left(\mathbf{z}_n \middle| \mathbf{x}_n, \theta^{old}\right)$$

$$= \frac{\sum_{\mathbf{z}_n} z_{nk} \prod_{k=1}^{3} N(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{z_{nk}} {\pi_\kappa}^{z_{nk}}}{\sum_{z_{nk}} \prod_{k=1}^{3} N(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{z_{nk}} {\pi_\kappa}^{z_{nk}}}$$

$$= \frac{\pi_k N(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{l=1}^{3} \pi_l N(\mathbf{x}_n|\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)}$$

Gaussian Mixture Models (Maximization Step)

$$\mathrm{G}(\theta) = E_{p(\boldsymbol{Z}|\boldsymbol{X},\theta)}[\ln p(\boldsymbol{X},\boldsymbol{Z}|\theta)]$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{3} E_{p(\boldsymbol{Z}|\boldsymbol{X},\theta)}[z_{nk}]\{\ln N(\boldsymbol{x}_n|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k) + \ln \pi_\kappa\}$$

$$= \sum_{n=1}^{N}\sum_{k=1}^{3} \gamma(z_{nk})\left\{-\frac{1}{2}(\boldsymbol{x}_n - \boldsymbol{\mu}_\kappa)^{\mathrm{T}}\boldsymbol{\Sigma}_\kappa{}^{-1}(\boldsymbol{x}_n - \boldsymbol{\mu}_\kappa) - \frac{1}{2}(F\ln 2\pi + \ln|\boldsymbol{\Sigma}_\kappa|) + \ln \pi_\kappa\right\}$$

$$\frac{dG(\theta)}{d\,\boldsymbol{\mu}_k} = 0 \qquad \frac{dG(\theta)}{d\,\boldsymbol{\Sigma}_k} = 0$$

Gaussian Mixture Models (Maximization Step)

$$\frac{dG(\theta)}{d\,\boldsymbol{\mu}_k} = \sum_{n=1}^{N} \gamma(z_{nk}) \Sigma_\kappa{}^{-1}(\boldsymbol{x}_n - \boldsymbol{\mu}_\kappa) = 0$$

$$\Rightarrow \mu_\kappa = \frac{\sum_{n=1}^{N} \gamma(z_{nk}) \boldsymbol{x}_{\boldsymbol{n}}}{\sum_{n=1}^{N} \gamma(z_{nk})}$$

$$\frac{dG(\theta)}{d\,\boldsymbol{\Sigma}_k} = \sum_{n=1}^{N} \gamma(z_{nk}) \{(\boldsymbol{x}_n - \boldsymbol{\mu}_\kappa)(\boldsymbol{x}_n - \boldsymbol{\mu}_\kappa)^{\mathrm{T}} - \boldsymbol{\Sigma}_k\} = 0$$

$$\Rightarrow \boldsymbol{\Sigma}_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk})(\boldsymbol{x}_n - \boldsymbol{\mu}_\kappa)(\boldsymbol{x}_n - \boldsymbol{\mu}_\kappa)^{\mathrm{T}}}{\sum_{n=1}^{N} \gamma(z_{nk})}$$

Gaussian Mixture Models (Maximization Step)

$$G(\theta) = \sum_{n=1}^{N}\sum_{k=1}^{3}\gamma(z_{nk})\left\{-\frac{1}{2}(\boldsymbol{x}_n - \boldsymbol{\mu}_\kappa)^{\mathrm{T}}\boldsymbol{\Sigma}_\kappa{}^{-1}(\boldsymbol{x}_n - \boldsymbol{\mu}_\kappa) - \frac{1}{2}(F\ln 2\pi + \ln|\boldsymbol{\Sigma}_\kappa|) + \ln\pi_\kappa\right\}$$

s.t. $\sum_{k=1}^{3}\pi_k = 1$

$$\mathrm{L}(\theta) = \mathrm{G}(\theta) - \lambda\left(\sum_{k=1}^{3}\pi_k - 1\right)$$

$$\frac{d\mathrm{L}(\theta)}{d\pi_k} = 0 \Rightarrow \quad \pi_\kappa = \frac{\sum_{n=1}^{N}\gamma(z_{nk})}{N}$$

Summary

Initialize

$\theta_x = \{\Sigma_1, \Sigma_2, \Sigma_3, \mu_1, \mu_2, \mu_3\}$

$\theta_z = \{\pi_1, \pi_2, \pi_3\}$

Expectation Step: $\gamma(z_{nk}) = \dfrac{\pi_k N(\boldsymbol{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{l=1}^{3} \pi_l N(\boldsymbol{x}_n|\boldsymbol{\mu}_l, \boldsymbol{\Sigma}_l)}$

Maximization Step:

$$\pi_\kappa = \frac{\sum_{n=1}^{N} \gamma(z_{nk})}{N} \quad \mu_\kappa = \frac{\sum_{n=1}^{N} \gamma(z_{nk}) x_n}{\sum_{n=1}^{N} \gamma(z_{nk})}$$

$$\Sigma_k = \frac{\sum_{n=1}^{N} \gamma(z_{nk})(\boldsymbol{x}_n - \boldsymbol{\mu}_\kappa)(\boldsymbol{x}_n - \boldsymbol{\mu}_\kappa)^{\mathrm{T}}}{\sum_{n=1}^{N} \gamma(z_{nk})}$$