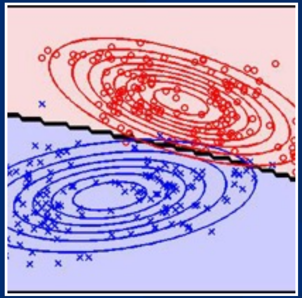
JOHNS HOPKINS
WHITING SCHOOL of ENGINEERING

Expectation Maximization

Mei-Ching Chen, PhD
Ben Rodriguez, PhD



Overview of Expectation Maximization

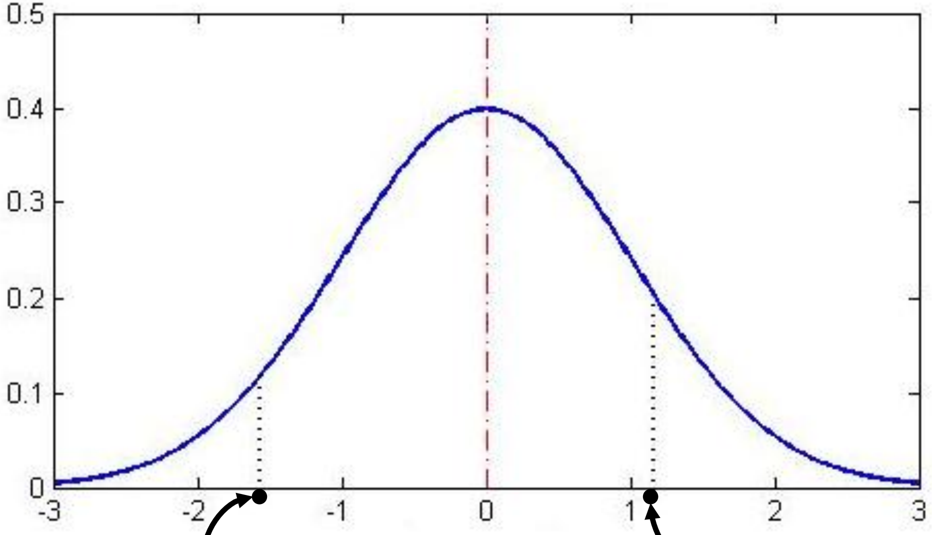
- Application
- The Purpose of EM
- EM Algorithm
- EM Examples
  - E-Step
    - Mixture of Gaussian Functions
    - Gaussian Normal
  - M-Step
    - Means
    - Standard Deviations
    - Mixing Probabilities

Applications of EM

- Application to Generated Data
- Application to Data Processing and Classification
  - Generated Feature Sets
  - Initial Findings
  - Sensitivity Analysis
- Future Research and Development

Why EM?

The idea behind the EM algorithm is that even though the data values of x_n are unknown the distribution $f(x_n|p)$ can be used to determine an estimate for p.



Probability of the data point being in the sub population

Probability of the data point being in the sub population

EM Algorithm

Given an initial estimate of $p_k^{(0)}, m_k^{(0)}, \sigma_k^{(0)}$, EM iterates the following computations until convergence to a local maximum of the likelihood function:

E-Step:

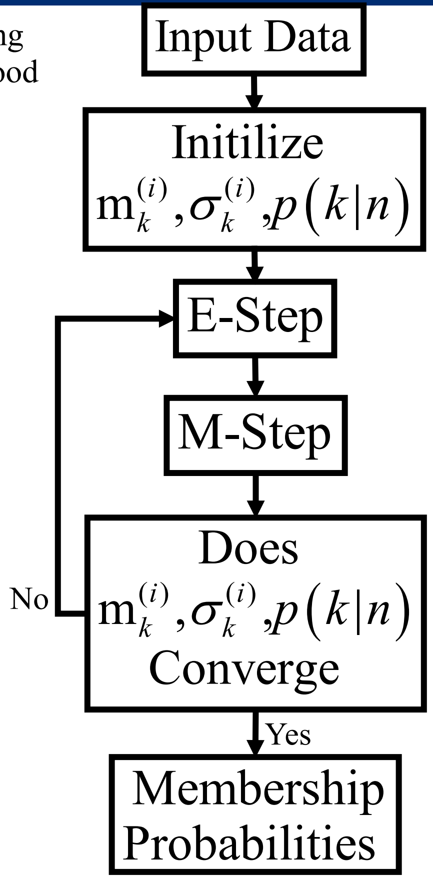
$$p^{(i)}(k|n) = \frac{p_k^{(i)} g\left(\mathbf{x}_n; \mathbf{m}_k^{(i)}, \sigma_\kappa^{(i)}\right)}{\sum_{k=1}^{K} p_k^{(i)} g\left(\mathbf{x}_n; \mathbf{m}_k^{(i)}, \sigma_\kappa^{(i)}\right)}$$

M-Step:

$$\mathbf{m}_k^{(i+1)} = \frac{\sum_{n=1}^{N} p^i(k|n) x_n}{\sum_{n=1}^{N} p^i(k|n)}$$

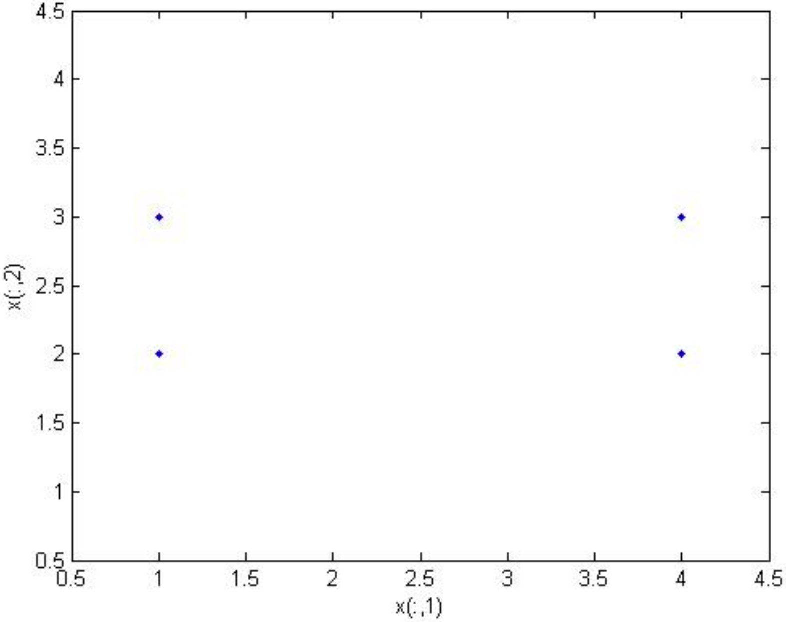
$$\sigma_k^{(i+1)} = \sqrt{\frac{1}{D} \frac{\sum_{n=1}^{N} p^i(k|n) \left\| \mathbf{x}_n - \mathbf{m}_k^{(i+1)} \right\|^2}{\sum_{n=1}^{N} p^i(k|n)}}$$

$$p_k^{(i+1)} = \frac{1}{N} \sum_{n=1}^{N} p^i(k|n)$$



Simple Example 1 of 8

$$x=\begin{bmatrix}1 & 2\\4 & 2\\1 & 3\\4 & 3\end{bmatrix}$$



Simple Example 2 of 8

Initial Parameters

$$x=\begin{bmatrix}1 & 2\\ 4 & 2\\ 1 & 3\\ 4 & 3\end{bmatrix},\quad \overline{x}_{col}=[2.5\quad 2.5]',\quad \sigma_{col}=[1.7321\quad 0.57735]',$$

$K \equiv$ number of sub populations

$$m=\overline{x}_{col}\,[1\quad 1]+\sigma_{col}\left[randn(1,K)\right]=\begin{bmatrix}2.5\\ 2.5\end{bmatrix}[1\quad 1]+\begin{bmatrix}1.7321\\ 0.57735\end{bmatrix}[-0.18671\quad 0.72579]$$

$$m=\begin{bmatrix}2.1766 & 3.7571\\ 2.3922 & 2.9190\end{bmatrix}$$

$$\sigma=\overline{\sigma}_{col}\,[1\quad 1]=[1.1547][1\quad 1]=[1.1547\quad 1.1547]$$

$$p(k|n)=\frac{[1\quad 1]}{K}=\begin{bmatrix}\frac{1}{2} & \frac{1}{2}\end{bmatrix}$$

E-Step

$$p^{(i)}(k|n)=\frac{p_k^{(i)} g\left(\mathbf{x}_n;\mathbf{m}_k^{(i)},\sigma_\kappa^{(i)}\right)}{\sum_{k=1}^{K} p_k^{(i)} g\left(\mathbf{x}_n;\mathbf{m}_k^{(i)},\sigma_\kappa^{(i)}\right)}$$

$$g\left(\mathbf{x}_n;\mathbf{m}_k^{(i)},\sigma_\kappa^{(i)}\right)=\frac{1}{\left(\sqrt{2\pi}\sigma_k\right)^D} e^{-\frac{1}{2}\left(\frac{\|\mathbf{x}_n-\mathbf{m}_k\|}{\sigma_k}\right)^2}$$

In the Expectation Step the conditional probability is generated. Computes the expected value of the x_n data using the current estimation of the parameter and the observed data.

Simple Example E-Step 3 of 8

$$g\left(\mathbf{x}_n;\mathbf{m}_k^{(i)},\sigma_\kappa^{(i)}\right)=\frac{1}{\left(\sqrt{2\pi}\sigma_k\right)^D}e^{-\frac{1}{2}\left(\frac{\|\mathbf{x}_n-\mathbf{m}_k\|}{\sigma_k}\right)^2}=\begin{bmatrix}0.0670 & 0.0324 & 0.0618 & 0.0299\\ 0.0050 & 0.0851 & 0.0069 & 0.1165\end{bmatrix}$$

$$x=\begin{bmatrix}1 & 2\\ 4 & 2\\ 1 & 3\\ 4 & 3\end{bmatrix},\quad m=\begin{bmatrix}2.1766 & 3.7571\\ 2.3922 & 2.9190\end{bmatrix},\quad \sigma=\begin{bmatrix}1.1547 & 1.1547\end{bmatrix}$$

Lets consider an example where $n=k=i=1$

$$g\left(\mathbf{x}_1;\mathbf{m}_1^{(1)},\sigma_1^{(1)}\right)=\frac{1}{\left(\sqrt{2\pi}\sigma_1\right)^2}e^{-\frac{1}{2}\left(\frac{\|\mathbf{x}_1-\mathbf{m}_1\|}{\sigma_1}\right)^2}=\frac{1}{\left(\sqrt{2\pi}1.1547\right)^2}e^{-\frac{1}{2}\left(\frac{\left\|\begin{bmatrix}1 & 2\end{bmatrix}\begin{bmatrix}\mathbf{-2.1766}\\ \mathbf{-2.3922}\end{bmatrix}\right\|^2}{1.1547^2}\right)}$$

$$=\frac{1}{8.3776}e^{-\frac{1}{2}\left(\frac{1.5382}{1.3333}\right)}=0.0670$$

Lets consider an example where only $k=1$

$$g\left(\mathbf{x}_n;\mathbf{m}_1^{(i)},\sigma_1^{(i)}\right)=\frac{1}{\left(\sqrt{2\pi}\sigma_1\right)^2}e^{-\frac{1}{2}\left(\frac{\|\mathbf{x}_n-\mathbf{m}_1\|}{\sigma_1}\right)^2}=\frac{1}{\left(\sqrt{2\pi}1.1547\right)^2}e^{-\frac{1}{2}\left(\frac{\left(\sum_{col}\left\|\begin{bmatrix}1 & 2\\ 4 & 2\\ 1 & 3\\ 4 & 3\end{bmatrix}-\begin{bmatrix}2.1766 & 2.3922\\ 2.1766 & 2.3922\\ 2.1766 & 2.3922\\ 2.1766 & 2.3922\end{bmatrix}\right\|^{Transpose}\right)^2}{1.1547^2}\right)}$$

$$=\frac{1}{8.3776}e^{-\frac{1}{2}\left(\frac{\left(\sum_{col}\begin{bmatrix}-1.1766 & 1.8234 & -1.1766 & 1.8234\\ -0.3922 & -0.3922 & 0.6078 & 0.6078\end{bmatrix}\right)^2}{1.3333}\right)}=\frac{\begin{bmatrix}0.5617 & 0.2713 & 0.5180 & 0.2502\end{bmatrix}}{8.3776}$$

$$=\begin{bmatrix}0.0670 & 0.0324 & 0.0618 & 0.0299\end{bmatrix}$$

Simple Example E-Step 4 of 8

$$g\left(\mathbf{x}_n;\mathbf{m}_k^{(i)},\sigma_\kappa^{(i)}\right)=\begin{bmatrix}0.0670 & 0.0324 & 0.0618 & 0.0299\\ 0.0050 & 0.0851 & 0.0069 & 0.1165\end{bmatrix}$$

$$p^{(i)}\left(k|n\right)=\frac{p_k^{(i)}g\left(\mathbf{x}_n;\mathbf{m}_k^{(i)},\sigma_\kappa^{(i)}\right)}{\sum_{k=1}^{K}p_k^{(i)}g\left(\mathbf{x}_n;\mathbf{m}_k^{(i)},\sigma_\kappa^{(i)}\right)}$$

$$=\frac{\begin{bmatrix}(0.5)0.0670 & (0.5)0.0324 & (0.5)0.0618 & (0.5)0.0299\\ (0.5)0.0050 & (0.5)0.0851 & (0.5)0.0069 & (0.5)0.1165\end{bmatrix}}{\begin{bmatrix}0.0360 & 0.0587 & 0.0344 & 0.0732\\ 0.0360 & 0.0587 & 0.0344 & 0.0732\end{bmatrix}}$$

$$=\frac{\begin{bmatrix}0.0335 & 0.0162 & 0.0309 & 0.0149\\ 0.0025 & 0.0425 & 0.0034 & 0.0582\end{bmatrix}}{\begin{bmatrix}0.0360 & 0.0587 & 0.0344 & 0.0732\\ 0.0360 & 0.0587 & 0.0344 & 0.0732\end{bmatrix}}$$

$$=\begin{bmatrix}0.9302 & 0.2758 & 0.8998 & 0.2041\\ 0.0693 & 0.7242 & 0.1002 & 0.7959\end{bmatrix}$$

M Step

The Maximization Step has three groups of equations:

- Means

$$\mathbf{m}_k^{(i+1)} = \frac{\sum_{n=1}^{N} p^i(k|n)\, x_n}{\sum_{n=1}^{N} p^i(k|n)}$$

- Standard Deviations

$$\sigma_k^{(i+1)} = \sqrt{\frac{1}{D} \frac{\sum_{n=1}^{N} p^i(k|n) \left\| \mathbf{x}_n - \mathbf{m}_k^{(i+1)} \right\|^2}{\sum_{n=1}^{N} p^i(k|n)}}$$

- Mixing Probabilities

$$p_k^{(i+1)} = \frac{1}{N} \sum_{n=1}^{N} p^i(k|n)$$

Simple Example M-Step 5 of 8

Means

$$\mathbf{m}_k^{(i+1)} = \frac{\sum_{n=1}^{N} p^i(k|n)x_n}{\sum_{n=1}^{N} p^i(k|n)} = \begin{bmatrix} 1.6232 & 3.6984 \\ 2.4779 & 2.5302 \end{bmatrix}$$

Let's consider when $k=1$

$$\mathbf{m}_1^{(i+1)} = \frac{\begin{bmatrix} (0.9302)1+(0.2758)4+(0.8998)1+(0.2041)4 \\ (0.9302)2+(0.2758)2+(0.8998)3+(0.2041)3 \end{bmatrix}}{2.31} = \frac{\begin{bmatrix} 3.7496 \\ 5.7239 \end{bmatrix}}{2.31} = \begin{bmatrix} 1.6232 \\ 2.4779 \end{bmatrix}$$

Let's consider when $k=2$

$$\mathbf{m}_2^{(i+1)} = \frac{\begin{bmatrix} (0.0698)1+(0.7242)4+(0.1002)1+(0.7959)4 \\ (0.0698)2+(0.7242)2+(0.1002)3+(0.7959)3 \end{bmatrix}}{1.69} = \frac{\begin{bmatrix} 6.2504 \\ 4.2761 \end{bmatrix}}{1.69} = \begin{bmatrix} 3.6984 \\ 2.5302 \end{bmatrix}$$

Simple Example M-Step 6 of 8

Standard Deviations

$$\sigma_k^{(i+1)} = \sqrt{\frac{1}{D}\frac{\sum_{n=1}^{N} p^i\left(k|n\right)\left\|\mathbf{x}_n - \mathbf{m}_k^{(i+1)}\right\|^2}{\sum_{n=1}^{N} p^i\left(k|n\right)}} = \left[0.9303 \quad 0.7290\right]$$

Let's consider when $k=1$

$$\sigma_1^{(i+1)} = \sqrt{\frac{\begin{pmatrix}(0.9302)(1-1.6232)^2 \\ +(0.9302)(2-2.4779)^2\end{pmatrix} + \begin{pmatrix}(0.2758)(4-1.6232)^2 \\ +(0.2758)(2-2.4779)^2\end{pmatrix} + \begin{pmatrix}(0.8998)(1-1.6232)^2 \\ +(0.8998)(3-2.4779)^2\end{pmatrix} + \begin{pmatrix}(0.2041)(4-1.6232)^2 \\ +(0.2041)(3-2.4779)^2\end{pmatrix}}{2(2.31)}}$$

$$= \sqrt{\frac{3.998}{4.62}} = 0.9303$$

Let's consider when $k=2$

$$\sigma_2^{(i+1)} = \sqrt{\frac{\begin{pmatrix}(0.0698)(1-3.6984)^2 \\ +(0.0698)(2-2.5302)^2\end{pmatrix} + \begin{pmatrix}(0.7242)(4-3.6984)^2 \\ +(0.7242)(2-2.5302)^2\end{pmatrix} + \begin{pmatrix}(0.1002)(1-3.6984)^2 \\ +(0.1002)(3-2.5302)^2\end{pmatrix} + \begin{pmatrix}(0.7959)(4-3.6984)^2 \\ +(0.7959)(3-2.5302)^2\end{pmatrix}}{2(1.69)}}$$

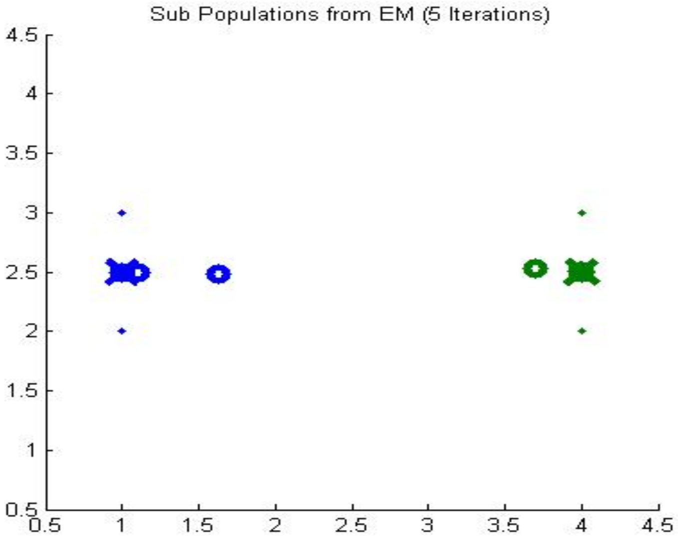
$$= \sqrt{\frac{1.7965}{3.38}} = 0.7290$$

Simple Example M-Step 7 of 8

Mixing Probabilities

$$p_k^{(i+1)} = \frac{1}{N}\sum_{n=1}^{N} p^i\left(k|n\right) = \begin{bmatrix} 0.9302+0.2758+0.8998+0.2041 \\ 0.0698+0.7242+0.1002+0.7959 \end{bmatrix} = \frac{1}{N}\begin{bmatrix} 2.31 \\ 1.69 \end{bmatrix} = \begin{bmatrix} 0.5775 \\ 0.4225 \end{bmatrix}$$

Simple Example 8 of 8



iter	$p^{(i)}(k \mid n)$	$m_k^{(i+1)}$	$\sigma_k^{(i+1)}$	$p_k^{(i+1)}$
1	$\begin{bmatrix} 0.9302 & 0.2758 & 0.8998 & 0.2041 \\ 0.0693 & 0.7242 & 0.1002 & 0.7959 \end{bmatrix}$	$\begin{bmatrix} 2.1766 & 3.7571 \\ 2.3922 & 2.9190 \end{bmatrix}$	$\begin{bmatrix} 1.1547 & 1.1547 \end{bmatrix}$	$\begin{bmatrix} 0.5775 & 0.4225 \end{bmatrix}$
2	$\begin{bmatrix} 0.9986 & 0.0384 & 0.9985 & 0.0355 \\ 0.0014 & 0.9616 & 0.0015 & 0.9645 \end{bmatrix}$	$\begin{bmatrix} 1.6232 & 3.6984 \\ 2.4779 & 2.5302 \end{bmatrix}$	$\begin{bmatrix} 0.9303 & 0.7290 \end{bmatrix}$	$\begin{bmatrix} 0.5177 & 0.4823 \end{bmatrix}$
3	$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1.1070 & 3.9955 \\ 2.4993 & 2.5008 \end{bmatrix}$	$\begin{bmatrix} 0.5890 & 0.3629 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$
4	$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1.0 & 4.0 \\ 2.5 & 2.5 \end{bmatrix}$	$\begin{bmatrix} 0.3536 & 0.3536 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$
5	$\begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1.0 & 4.0 \\ 2.5 & 2.5 \end{bmatrix}$	$\begin{bmatrix} 0.3536 & 0.3536 \end{bmatrix}$	$\begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$

So What

- Why not k-mean?
  - There is a firm assignment of samples to the assigned cluster
  - EM assigns samples based on posterior probabilites
  - K-means is not a classifier

- Applying mixture models will give better results for the mean, covariance, and probabilities

- Using the Bayes classifier with EM will allow for class labels

References

- Tomasi, C., Estimating Gaussian Mixture Densities with EM – A Totorial

- Dempster, A.P., Laird, N.M., and Rubin, D.B., 1977, Maximum Likelihood from Incomplete Data via the EM Algorithm, *Journal of the Royal Statistical Society, Series B (Methodological)*, Volume 39, Issue 1, (1977), 1-38. DemLaiRub77.pdf

- Moon, T. K., The Expectation Maximization Algorithm, *IEEE Signal Processing Magazine*, November 1996, 47-60. Moon_EMArticle.pdf

- Duda, R. O., Hart, P.E., Stork, D. G., <u>Patern Classification</u>, Second Edition