Expectation Maximization Algorithm

Ahmet Ademoglu, *PhD*
Bogazici University
Institute of Biomedical Engineering

EM Algorithm

Say that the probability of the temperature outside your window for each of the 24 hours of a day $x \in R^{24}$ depends on the season $\theta \in \{$ summer, fall, winter, spring$\}$, and that you know the seasonal temperature distribution $p(x|\theta)$. But say you can only measure the average temperature $y = T(x)$ for the day, and you'd like to guess what season θ it is (for example, is spring here yet?). The maximum likelihood estimate of θ maximizes $p(y|\theta)$, but in some cases this may be hard to find. That's when EM is useful it takes your observed data y, iteratively makes guesses about the complete data x, and iteratively finds the that maximizes $p(x|\theta)$ over θ. In this way, EM tries to find the maximum likelihood estimate of θ given y.

Ector's Problem

Let the random variable X_1, represent the number of round dark objects, X_2, represent the number of square dark objects, and X_3, represent the number of light objects.
Let $\mathbf{x} = [x_1, x_2, x_3]^T$ be the vector of values the random variables take for some image.

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = \left(\frac{n!}{x_1!x_2!x_3!}\right)\left(\frac{1}{4}\right)^{x_1}\left(\frac{1}{4}+\frac{p}{4}\right)^{x_2}\left(\frac{1}{2}-\frac{p}{4}\right)^{x_3}$$

where p is an unkown parameter and $n = x_1 + x_2 + x_3$.

Ector's Problem

Let $\mathbf{y} = [y_1, y_2]^T$ be the number of dark objects and number of light objects detected, respectively, so that $y_1 = x_1 + x_2$ and $y_2 = x_3$ and let the corresponding random variables be Y_1, and Y_2. The likelihood is

$$P(Y_1 = y_1|p) = \left(\frac{n}{y_1}\right)\left(\frac{1}{2} + \frac{p}{4}\right)^{y_1}\left(\frac{1}{2} - \frac{p}{4}\right)^{y_2}$$

Expectation Step

We assume the latent variables x_1 and x_2 and compute their conditional expectations;

$x_1^{k+1} = E[x_1|y_1, p^k]$ and $x_1^{k+1} = E[x_2|y_2, p^k]$.

x_3 is directly observed since $x_3 = y_3$.

Ector's Problem

If (X_1, X_2, X_3) has a multinomial distribution with probabilities (p_1, p_2, p_3) then

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = \left(\frac{n!}{x_1!x_2!x_3!}\right) p_1^{x_1} p_2^{x_2} p_3^{x_3}$$

$$\begin{aligned}
P(X_1 + X_2 = y_1, X_3 = x_3) &= \sum_{i=0}^{y_1} P(X_1 = i, X_2 = y_1 - i, X_3 = x_3) \\
&= \frac{(y_1 + x_3)!}{y_1!x_3!} p_3^{x_3} \sum_{i=0}^{y_1} \frac{y_1!}{(y_1 - i)!} p_1^i p_2^{y_1 - i} \\
&= \frac{n!}{y_1!x_3!} (p_1 + p_2)^{y_1} p_3^{x_3} \\
P(X_1 + X_2 = y_1, X_3 = x_3 | \mathbf{p^k}) &= P(Y_1 = y_1, Y_2 = y_2 | \mathbf{p^k})
\end{aligned}$$

Ector's Problem

To compute $x_1^{k+1} = E[x_1|y_1, y_2, p^k]$ we first determine

$$P(X_1 = x_1|Y_1 = y_1, Y_2 = y_2, \mathbf{p}^k) = \frac{P(X_1 = x_1, Y_1 = y_1, Y_2 = y_2|\mathbf{p}^k)}{P(Y_1 = y_1, Y_2 = y_2|\mathbf{p}^k)}$$

$$P(X_1 = x_1, Y_1 = y_1, Y_2 = y_2|\mathbf{p}^k) = P(X_1 = x_1, X_2 = x_2, X_3 = x_3)$$

$$= \frac{P(X_1 = x_1, X_2 = y_1 - x_1, X_3 = x_3|\mathbf{p}^k)}{P(Y_1 = y_1, Y_2 = y_2|\mathbf{p}^k)}$$

$$P(X_1 = x_1|Y_1 = y_1, Y_2 = y_2, p^k) = \frac{y_1!}{x_1!(y_1 - x_1)!} p_1^{x_1} p_2^{y_1 - x_1} \frac{1}{(p_1 + p_2)^{y_1}}$$

Ector's Problem

For computing $x_1^{k+1} = E[x_1|y_1, y_2, p^k]$
we can use

$$P(X_1 = x_1|Y_1 = y_1, Y_2 = y_2, \mathbf{p}^k)$$

$$= \sum_{x_1=0}^{y_1} x_1 \frac{y_1!}{x_1!(y_1 - x_1)!} p_1^{x_1} p_2^{y_1 - x_1} \frac{1}{(p_1 + p_2)^{y_1}}$$

$$E[x_1|y_1, y_2, p^k] = y_1 \frac{p_1}{p_1 + p_2}$$

Likewise, $x_2^{k+1} = E[x_2|y_1, y_2, p^k]$ can be determined as

$$y_1 \frac{p_2}{p_1 + p_2}$$

Ector's Problem

Maximization Step

We maximize the log-likelihood with respect to the unknown parameter p,

$$\frac{d}{dp}\log P(X_1 = x1, X_2 = x_2, X_3 = x_3) =$$

$$\frac{d}{dp}\log\left(\frac{n!}{x_1!x_2!x_3!}\right)\left(\frac{1}{4}\right)^{x_1}\left(\frac{1}{4}+\frac{p}{4}\right)^{x_2}\left(\frac{1}{2}-\frac{p}{4}\right)^{x_3} = 0$$

which yields

$$p\left(^{k+1}\right) = \frac{2x_2^k - x_3}{x_2^k + x_3}$$

EM Algorithm

To use EM, you must be given some observed data y, a parametric density $p(y|\theta)$, a description of some complete data x that you wish you had, and the parametric density $p(x|\theta)$. x can be modeled as a continuous random variable X with density $p(x|\theta)$, where $\theta \in \Theta$. You do not observe X directly; instead, you observe a realization y of the random variable $Y = T(X)$ for some function T.

EM Algorithm

$$\hat{\theta}_{MLE} = \arg\max_{\theta \in \Theta} \log p(y|\theta)$$

Step 1 Pick an initial guess θ^0.

Step 2 Given the observed data y calculate how likely it is that the complete data is exactly x, that is, the conditional distribution $p(x|y, \theta^m)$.

Step 3 Make a new guess of θ that maximizes (the expected) $\log p(x|y, \theta^m)$ by integrating over all possible values of x.

$$Q(\theta|\theta^m) = E_{X|y,\theta^m}[\log p(X|\theta)] = \int \log p(x|\theta) p(x|y, \theta^m) dx$$

Step 4 Repeat 2 to 3 until convergence.

$$Q(\theta|\theta^m) = \int \log p(x|\theta) \frac{p(x|\theta^m)}{p(y|\theta^m)} dx$$

Toy Example

n kids choose a toy out of four choices with histogram $Y = [Y_1 \cdots Y_4]^T$ where Y is the number of kids that chose toy 1, etc.

Y : distributed according to a multinomial distribution

$$P(y|\theta) = \frac{n!}{y_1! y_2! y_3! y_4!} p_1^{y_1} p_2^{y_2} p_3^{y_3} p_4^{y_4}$$

with $p \in (0,1)^4$ and $p_1 + p_2 + p_3 + p_4 = 1$.
$p_\theta = [\frac{1}{2} + \frac{1}{4}\theta \quad \frac{1}{4}(1-\theta) \quad \frac{1}{4}(1-\theta) \quad \frac{1}{4}\theta]$, $\theta \in (0,1)$.

$$P(y|\theta) = \frac{n!}{y_1! y_2! y_3! y_4!} \left(\frac{1}{2} + \frac{1}{4}\theta\right)^{y_1} \left(\frac{1-\theta}{4}\right)^{y_2} \left(\frac{1-\theta}{4}\right)^{y_3} \left(\frac{\theta}{4}\right)^{y_4}$$

Toy Example

The complete data $X = [X_1 \cdots X_5]^T$ has a multinomial distribution with number of trials n and the probability of $q_\theta = [\frac{1}{2} \quad \frac{1}{4}\theta \quad \frac{1}{4}(1-\theta) \quad \frac{1}{4}(1-\theta) \quad \frac{1}{4}\theta]$, $\theta \in (0,1)$.

$$Y = T(X) = X = [X_1 + X_2 \quad X_3 \quad X_4 \quad X_5]^T$$

Then the likelihood of a realization x of the complete data is $P(x|\theta) = \frac{n!}{x_1!x_2x_3!x_4!x_5!}\left(\frac{1}{2}\right)^{x_1}\left(\frac{\theta}{4}\right)^{x_2+x_5}\left(\frac{1-\theta}{4}\right)^{x_3+x_4}$

For EM, we must maximize the Q-function:

$$\theta^{m+1} = \underset{\theta\in(0,1)}{\arg\max} Q(\theta|\theta^m) = \underset{\theta\in(0,1)}{\arg\max} E_{X|y,\theta^m}[\log p(X|\theta)]$$

Toy Example

The derivative only applies to θ dependent terms so

$$\theta^m = \underset{\theta\in(0,1)}{\arg\max} E_{X|y,\theta^m}[(X_2 + X_5)\log\theta + (X_3 + X_4)\log(1-\theta)]$$

$$= \underset{\theta\in(0,1)}{\arg\max}[\log\theta(E_{X|y,\theta^m}[X_2] + E_{X|y,\theta^m}[X_5])+$$
$$(1 - \log\theta)(E_{X|y,\theta^m}[X_3] + E_{X|y,\theta^m}[X_4])]$$

$$P(x|y,\theta) = \tfrac{y_1!}{x_1!x_2!}\left(\tfrac{2}{2+\theta}\right)^{x_!}\left(\tfrac{\theta}{2+\theta}\right)^{x_2} 1_{\{x_1+x_2=y_1\}} \textstyle\prod_{i=3}^{5} 1_{\{x_i=y_{i-1}\}}$$

$$E_{X|y,\theta}[X] = [\frac{2}{2+\theta}y_1 \quad \frac{\theta}{2+\theta}y_1 \quad y_2 \quad y_3 \quad y_4]^T$$

$$\theta^{m+1} = \underset{\theta\in(0,1)}{\arg\max}\left(\log\theta\left(\tfrac{\theta^m y_1}{2+\theta^m} + y_4\right) + \log(1-\theta)(y_2+y_3)\right)$$

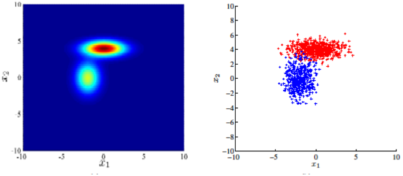
$$= \frac{\frac{\theta^m y_1}{2+\theta^m}+y_4}{\frac{\theta^m y_1}{2+\theta^m}+y_2+y_3+y_4}$$

Gaussian Mixture Model

Now given n i.i.d. samples $y_1, y_2, \ldots y_n \in \mathcal{R}^d$ from a GMM with k components, the estimate its parameter set $\theta = \{(w_j, \mu_j, \Sigma_j)\}_{j=1}^{k}$.

$$\phi(y|\mu, \Sigma) =\triangleq \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2} exp\Big(-\frac{1}{2}(y-\mu)^T\Sigma^{-1}(y-\mu)\Big)}$$

$$p(y|\theta) = \sum_{i=1}^{k} w_i\phi(y|\mu_i, \Sigma_i)$$



EM clustering by a Gaussian Mixture Model

$$p(y|\theta) = \sum_{i=1}^{k} w_i \phi(\mu_i, \Sigma_i) = \sum_{i=1}^{k} w_i \frac{\exp(-\frac{1}{2}(y-\mu_j)^T \Sigma_j^{-1}(y-\mu_j))}{(2\pi)^{d/2}|\Sigma|^{1/2}}$$

Consider for simplicity just one random observation Y from the GMM. The complete data be $X = (Y, Z)$, where $Z \in \{1, \ldots, k\}$ is is a discrete random variable that defines which Gaussian component the data Y came from, so $P(Z = i) = w_i$, $i = 1, \ldots, k$, and $(Y|Z = i) \sim \mathcal{N}_d(\mu_i, \Sigma_i)$, $i = 1, \ldots, k$. The density of the complete data X is

$$p_X(Y = y, Z = i|\theta) = w_i \frac{\exp(-\frac{1}{2}(y-\mu_j)^T \Sigma_j^{-1}(y-\mu_j))}{(2\pi)^{d/2}|\Sigma|^{1/2}}$$

$$p(y|\theta) = \sum_{i=1}^{k} w_i p_X(Y = y, Z = i|\theta)$$

Gaussian Mixture Model

Let the complete data X be the observed data Y plus some missing (also called latent or hidden) data$Z$, so that $X = (Y, Z)$. The Q-function over the domain of Z because the only random part of X is Z is

$$
\begin{aligned}
Q(\theta|\theta_m) &= E_{X|y,\theta^m}[\log p_X(X|\theta)] \\
&= E_{Z|y,\theta^m}[\log p_X(y, Z|\theta)] \\
&= \int_{\mathcal{Z}} \log p_X(y, z|\theta) p_{Z|Y}(z|y, \theta^m) dz
\end{aligned}
$$

Gaussian Mixture Model

Define γ_{ij}^m to be your guess at the m^{th} iteration of the probability that the i^{th} sample belongs to the j^{th} Gaussian component, that is,

$$\gamma_{ij}^m \triangleq P(Z_i = j | Y_i = y_i, \theta^m) = \frac{w_j^m \phi(y_i | \mu_j^m, \Sigma_j^m)}{\sum_{l=1}^{k} w_l^m \phi(y_i | \mu_l^m, \Sigma_l^m)}$$

which satisfies $\sum_{j=1}^{k} \gamma_{ij}^m = 1$.

Gaussian Mixture Model

The E-step:

$$Q_i(\theta|\theta^m) = E_{Z_i|y_i,\theta^m}[\log p_X(y_i, j|\theta)] = \sum_{j=1}^{k} \gamma_{ij}^m \log p_X(y_i, j|\theta)]$$

$$= \sum_{j=1}^{k} \gamma_{ij}^m \log w_j \phi(y_i|\mu_j^m, \Sigma_j)$$

$$= \sum_{j=1}^{k} \gamma_{ij}^m \Big(\log w_j - \frac{1}{2} \log|\Sigma_j| - \frac{1}{2}(y_i \mu_j)^T \Sigma_j^{-1}(y_i - \mu_j)\Big) + constant$$

$$Q(\theta|\theta^m) = \sum_{i=1}^{n} Q_i(\theta|\theta^m)$$

Gaussian Mixture Model

Let $n_j^m = \sum_{i=1}^n \gamma_{ij}^m$ then $Q(\theta|\theta^m)$

$$= \sum_{j=1}^{k} n_j^m \Big(\log w_j - \tfrac{1}{2} \log |\Sigma_j| \Big) - \tfrac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{k} \gamma_{ij}^m (y_i - \mu_j)^T \Sigma_j^{-1} (y_i - \mu_j)$$

M-Step is to $\underset{\theta}{maximize}\ Q(\theta|\theta^m)$

subject to $\sum_{j=1}^{k} w_j = 1,\ w_j \geq 0,\ j = 1, \ldots, k.$

If we form the Lagrangian,

$$\mathcal{L}(\theta, \lambda) = \sum_{j=1}^{k} n_j^m \Big(\log w_j - \tfrac{1}{2} \log |\Sigma_j| \Big)$$

$$-\tfrac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{k} \gamma_{ij}^m (y_i - \mu_j)^T \Sigma_j^{-1} (y_i - \mu_j) + \lambda \Big(\sum_{j=1}^{k} w_j - 1 \Big)$$

Gaussian Mixture Model

$$\frac{\partial \mathcal{L}}{\partial w_l} = \frac{\partial}{\partial w_l}\Big(\sum_{j=1}^{k} n_j^m \log w_j + \lambda\Big(\sum_{j=1}^{k} w_j - 1\Big)\Big) = 0, \;\; l = 1, \ldots, k.$$

which yields

$$w_j^{m+1} = \frac{n_j^m}{\sum\limits_{j=1}^{k} n_j^m} = \frac{n_j^m}{n}, \quad j = 1, \ldots, k.$$

Similarly,

$$\frac{\partial \mathcal{L}}{\partial \mu_j} = \Sigma_j^{-1}\Big(\sum_{i=1}^{n} \gamma_{ij}^m y_i - n_j^m \mu_j\Big) = 0, \quad j = 1, \ldots, k.$$

$$\mu_j^{m+1} = \frac{1}{n_j^m}\sum_{i=1}^{n} \gamma_{ij}^m y_i, \quad j = 1, \ldots, k.$$

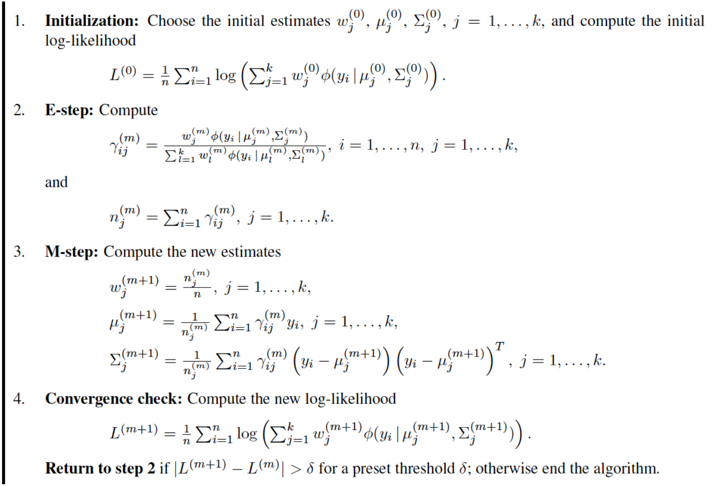
Gaussian Mixture Model

$$\frac{\partial \mathcal{L}}{\partial \Sigma_j} = -\frac{1}{2} n_j^m \frac{\partial}{\partial \Sigma_j} \log |\Sigma_j| - \frac{1}{2} \sum_{i=1}^{n} \gamma_{ij}^m \frac{\partial}{\partial \Sigma_j} (y_i - \mu_j)^T \Sigma_j^{-1} (y_i - \mu_j)$$

$= 0, \quad j = 1, \ldots, k$
we get

$$\Sigma_j^{m+1} = \frac{1}{n_j^m} \sum_{i=1}^{n} \gamma_{ij}^m \frac{\partial}{\partial \Sigma_j} (y_i - \mu_j^{m+1})(y_i - \mu_j^{m+1})^T, \quad j = 1, \ldots, k.$$

EM algorithm for estimating GMM parameters



1. **Initialization:** Choose the initial estimates $w_j^{(0)}$, $\mu_j^{(0)}$, $\Sigma_j^{(0)}$, $j = 1, \ldots, k$, and compute the initial log-likelihood

$$L^{(0)} = \frac{1}{n} \sum_{i=1}^{n} \log \left(\sum_{j=1}^{k} w_j^{(0)} \phi(y_i \,|\, \mu_j^{(0)}, \Sigma_j^{(0)}) \right).$$

2. **E-step:** Compute

$$\gamma_{ij}^{(m)} = \frac{w_j^{(m)} \phi(y_i \,|\, \mu_j^{(m)}, \Sigma_j^{(m)})}{\sum_{l=1}^{k} w_l^{(m)} \phi(y_i \,|\, \mu_l^{(m)}, \Sigma_l^{(m)})}, \; i = 1, \ldots, n, \; j = 1, \ldots, k,$$

and

$$n_j^{(m)} = \sum_{i=1}^{n} \gamma_{ij}^{(m)}, \; j = 1, \ldots, k.$$

3. **M-step:** Compute the new estimates

$$w_j^{(m+1)} = \frac{n_j^{(m)}}{n}, \; j = 1, \ldots, k,$$

$$\mu_j^{(m+1)} = \frac{1}{n_j^{(m)}} \sum_{i=1}^{n} \gamma_{ij}^{(m)} y_i, \; j = 1, \ldots, k,$$

$$\Sigma_j^{(m+1)} = \frac{1}{n_j^{(m)}} \sum_{i=1}^{n} \gamma_{ij}^{(m)} \left(y_i - \mu_j^{(m+1)} \right) \left(y_i - \mu_j^{(m+1)} \right)^T, \; j = 1, \ldots, k.$$

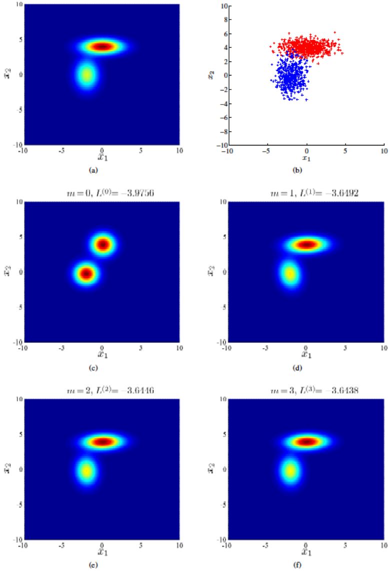
4. **Convergence check:** Compute the new log-likelihood

$$L^{(m+1)} = \frac{1}{n} \sum_{i=1}^{n} \log \left(\sum_{j=1}^{k} w_j^{(m+1)} \phi(y_i \,|\, \mu_j^{(m+1)}, \Sigma_j^{(m+1)}) \right).$$

Return to step 2 if $|L^{(m+1)} - L^{(m)}| > \delta$ for a preset threshold δ; otherwise end the algorithm.

A GMM FITTING EXAMPLE

$\mu_1 = \begin{bmatrix} 0 \\ 4 \end{bmatrix}$, $\mu_2 = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$, $\Sigma_1 = \begin{bmatrix} 3 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$, $\Sigma_2 = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$,
$w_1 = 0.6,\ w_2 = 0.4$
Initial values are $\mu_1^0 = \begin{bmatrix} 0.0823 \\ 3.9189 \end{bmatrix}$,$\mu_2^0 = \begin{bmatrix} -2.0706 \\ -2.2327 \end{bmatrix}$,
$\Sigma_1^0 = \Sigma_2^0 = I_2,\ w_1^0 = w_2^0 = 0.5$ and $\delta = 10^{-3}$.
After 3 iterations
$\mu_1^3 = \begin{bmatrix} 0.0806 \\ 3.9445 \end{bmatrix}$, $\mu_2^3 = \begin{bmatrix} -2.0181 \\ -0.1740 \end{bmatrix}$,
$\Sigma_1^3 = \begin{bmatrix} 2.7452 & 0.0568 \\ 0.0568 & 0.4821 \end{bmatrix}$, $\Sigma_2^3 = \begin{bmatrix} 0.8750 & -0.0153 \\ -0.0153 & 1.7935 \end{bmatrix}$,
$w_1^3 = 0.5966,\ w_2^3 = 0.4034$.



(a)

(b)

$m=0$, $L^{(0)} = -3.9756$

$m=1$, $L^{(1)} = -3.6492$

(c)

(d)

$m=2$, $L^{(2)} = -3.6446$

$m=3$, $L^{(3)} = -3.6438$

(e)

(f)

Hidden Markov Model

A sequence $Y = [Y_1 \; Y_2, \ldots, Y_T]$, where $Y_t \in \mathcal{R}^d$, $t = 1, \ldots, T$. The complete data: $X = (Y, Z)$, the observed sequence Y plus the (hidden) sequence Z: $Z = [Z1 \, Z_2, \ldots, Z_T]$ where $Z_t \in \{1, 2, \ldots, k\}, t = 1, \ldots, T$.

In genomics one might be modeling a DNA sequence as an HMM, where the hidden state values are coding region or non-coding region. Thus each $Z_t \in \{coding, non - coding\}$, $k = 2$, and each observation is $Y_t \in \{A, T, C, G\}$.

$$p(x) = p(y, z) = \prod_{\tau=1}^{T} p(y_\tau | z_\tau) P(z_1) \prod_{\tau=2}^{T} P(z_t | z_{t-1})$$

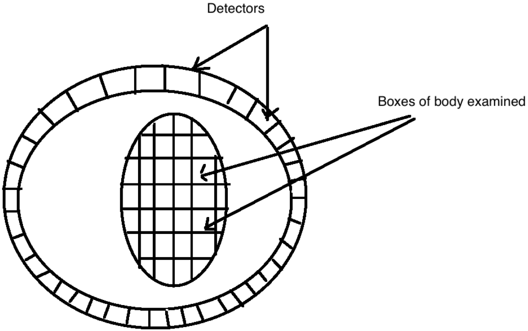
Initial probability distribution:$\pi = [\pi_1, \ldots, \pi_k]^T$, $\pi_i = P(Z_1 = i)$. A transition probability matrix $\mathbf{P} \in \mathcal{R}^{k \times k}$ specifies the probability of transitioning from state i to j: $P_{ij} = P(Z_t = j | Z_{t-1} = i)$,

Hidden Markov Model

The probability distribution of observations $Y \in \mathcal{R}^d$ given hidden state i; $p(Y_t = y|Z_t = i) = p(y|\theta_i)$.
In modeling a DNA sequence, the $Z_t = i$ is a pmf parameter that specifies the probabilities of A, T, C, and G being observed if the hidden state is $Z_t = i$.
Then the parameter set θ_i for the i^{th} hidden state is $\theta_i\{(w_{ij}, \mu_{ij,\Sigma_{ij}})\}_{j=1}^{M_i}$ where M_i is the number of components for the GMMs of i^{th} hidden state.
Then the total parameter set to estimate for HMM :
$\theta = \{\pi, \mathbf{P}, \theta_1, \ldots, \theta_k\}$

Emission Tomography (PET SPECT)

The probability of detecting an event originated from box j in detector tube i:
P(event detected in tube i —event occurred in box j) $= H_{ij}$,
where $\mathbf{H}$ is called the system matrix.



Emission Tomography (PET SPECT)

The average/expected number of events detected in tube i is then $E[g_i] = \sum_j H_{ij} f_j$.

In matrix-vector notation:

$$E[\mathbf{g}] = \mathbf{Hf}$$

Image reconstruction is to invert this equation to solve for the image $\mathbf{f}$.

If we could observe G directly, the solution to the whole problem would be simple: $\hat{f}_i = \sum_i G_{ij}$

G_{ij} : Poisson distributed complete data (observed + hidden) and The likelihood function is

$$L(\mathbf{f}) = P(\mathbf{G}|\mathbf{f}) = \prod_i \prod_j P(G_{ij}|\mathbf{f}) = \prod_i \prod_j e^{E[G_{ij}]} \frac{E[G_{ij}]^{G_{ij}}}{G_{ij}!}$$

Emission Tomography (PET SPECT)

$E[G_{ij}]$ is the expected number of emissions from voxel j measured in tube i:

$$E[G_{ij}] = f_j H_{ij}$$

Sum of Poisson Processes is another Poisson :
Given two Poisson distributed random variables X_1 and X_2 and their corresponding distribution parameters λ_1 and λ_2. If $Y = X_1 + X_2$, then the probability density function of Y is

$$P(Y=n) = P(X_1+X_2=n) = \sum_{k=0}^{n} P(X_1=k)P(X_2=n-k)$$

$$= \sum_{k=0}^{n} e^{-\lambda_1}\tfrac{\lambda_1^k}{k!} e^{-\lambda_2}\tfrac{\lambda_2^k}{(n-k)!} = \tfrac{1}{n!} e^{-(\lambda_1+\lambda_2)} \sum_{k=0}^{n} \tfrac{n!}{k!(n-k)!}\lambda_1^k \lambda_2^{n-k}$$

$P(Y=n) = \frac{1}{n!} e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n$ which is another Poisson with parameter $\sum_i \lambda_i$.

Emission Tomography (PET SPECT)

Conditional Expectation of a Poisson Process is Binomial :

$$E[X_1 = x | X_1 + X_2 = y] = \frac{P(X_1 = x, X_1 + X_2 = y)}{P(X_1 + X_2 = y)}$$

$$P(X_1 = x, X_1 + X_2 = y) = P(X1 = x)P(X2 = y - x)$$

$$= e^{-\lambda_1}\frac{\lambda_1^k}{x!}e^{-\lambda_2}\frac{\lambda_2^{y-x}}{(y-x)!}$$

$$P(X_1 + X_2 = y) = e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1 + \lambda_2)^y}{y!}$$

$$E[X_1 = x | X_1 + X_2 = y] = \frac{y!}{x!(y-x)!}\frac{e^{-\lambda_1}e^{-\lambda_2}}{e^{-(\lambda_1+\lambda_2)}}\frac{\lambda_1^x\lambda_2^{y-x}}{(\lambda_1 + \lambda_2)^y}$$

which shows that it is Binomially distributed (n, p) with parameters $(y = x1 + x2, \frac{\lambda_1}{\lambda_1+\lambda_2})$.

Emission Tomography (PET SPECT)

The log-likelihood function is

$$\log L(\mathbf{f}) = \sum_i \sum_j -f_j H_{ij} + G_{ij} \ln f_j H_{ij} - \ln G_{ij}!$$

Expectation :

$$E\left[\sum_i \sum_j -f_j H_{ij} + G_{ij} \ln f_j H_{ij} - \ln G_{ij}! | \mathbf{g}, \hat{\mathbf{f}}^{\mathbf{k}}\right]$$

$$\sum_i \sum_j \Big(- f_j H_{ij} + E[G_{ij} | \mathbf{g}, \hat{\mathbf{f}}^{\mathbf{k}}] \ln f_j H_{ij} - E[G_{ij}! | \mathbf{g}, \hat{\mathbf{f}}^{\mathbf{k}}] \Big)$$

Emission Tomography (PET SPECT)

Considering the constraint that $g_i = \sum_j G_{ij}$ and G_{ij} are independent Poisson random variables, the conditional probability of G^{ij} upon g_i

$$E[G_{ij}|\mathbf{g}, \hat{\mathbf{f}}^{\mathbf{k}}]$$

is Binomially distributed with parameters $(\sum_j G_{ij}, \frac{E[G_{ij}]}{\sum_j E[G_{ij}]})$.

Since $E[G_{ij}] = f_j H_{ij}$, the expected value of the binomial distribution is

$$E[G_{ij}|\mathbf{g}, \hat{\mathbf{f}}^{\mathbf{k}}] = g_i \frac{\hat{f}_j^k H_{ij}}{\sum_m \hat{f}_m^k H_{im}}$$

Emission Tomography (PET SPECT)

Maximization :

$$\frac{\partial}{\partial f_l} E[\log L(\mathbf{f})|\mathbf{g}, \hat{\mathbf{f}}^k] = 0 = -\sum_i H_{il} + \sum_i E[G_{il}|\mathbf{g}, \hat{\mathbf{f}}^k] \frac{1}{f_l}$$

$$f_l = f^{k+1} = \frac{\sum_i E[G_{il}|\mathbf{g}, \hat{\mathbf{f}}^k]}{\sum_i H_{il}} = \frac{\hat{f}_l^k}{\sum_i H_{il}} \sum_i \frac{H_{il} g_i}{\sum_m \hat{f}_m^k H_{im}}$$

EM ON HIDDEN MARKOV MODELS

The states : $\{1, 2, \ldots, K\}$
Joint distribution for a sequence of N observations under this model is

$$p(\mathbf{x}_1, \ldots \mathbf{x}_N) = \prod_{n=1}^{N} p(\mathbf{x}_n | \mathbf{x}_1, \ldots, \mathbf{x}_{n-1})$$

For a first order HMM

$$p(\mathbf{x}_1, \ldots \mathbf{x}_N) = p(\mathbf{x}_1) \prod_{n=2}^{N} p(\mathbf{x}_n | \mathbf{x}_{n-1})$$

Using the latent variables $\mathbf{z}_n$, the joint distribution is

$$p(\mathbf{x}_1, \ldots \mathbf{x}_N, \mathbf{z}_1, \ldots \mathbf{z}_N) = p(\mathbf{z_1}) \prod_{n=2}^{N} p(\mathbf{z}_n | \mathbf{z}_{n-1}) \prod_{n=1}^{N} p(\mathbf{x}_n | \mathbf{z}_n)$$

EM ON HIDDEN MARKOV MODELS

$\mathbf{z}_n$: discrete multinomial variables describing which component of the mixture is responsible for generating the corresponding observation $\mathbf{x}_n$.

$\mathbf{z}_n$ depends on the state of the previous latent variable $\mathbf{z}_{n-1}$ with probability $p(\mathbf{z}_n|\mathbf{z}_{n-1})$ and is denoted by

$\mathbf{A}$ whose entries $a_{jk} = P(z_{nk} = 1|z_{n-1j} = 1)$ are called transition probabilities.

$$p(\mathbf{z}_n|\mathbf{z}_{n-1\mathbf{A}}) = \prod_{k=1}^{K} \prod_{j=1}^{K} A_{jk}^{z_{n-1j} z_{nk}}$$

The probability of the $\mathbf{z}_1$: initial latent node is

$$p(\mathbf{z}_1|\pi) = \prod_{k=1}^{K} \pi_k^{z_{1k}}$$

set by π with $\pi_k = p(z_{1k} = 1)$ and $\sum_{k=1}^{K} \pi_k = 1$

EM ON HIDDEN MARKOV MODELS

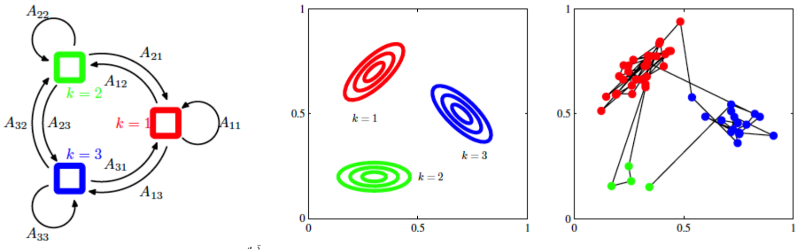
The probabilities of the observations: alse called emission probabilities are $p(\mathbf{x}_n|\mathbf{z}_n, \phi) = \prod_{k=1}^{K} p(\mathbf{x}_n|\phi_k)^{z_{nk}}$ where $\mathbf{z}_n$ determines from which node the observation has been sampled from.

The joint distribution is

$$p(\mathbf{X}, \mathbf{Z}|\theta) = p(\mathbf{z}_1|\pi)\left[\prod_{n=2}^{N} p(\mathbf{z}_n|\mathbf{z}_{n-1}), \mathbf{A}\right] \prod_{m=1}^{N} p(\mathbf{x}_m|\mathbf{z}_m, \phi)$$

HMM is defined by $\theta = \{\mathbf{A}, \pi, \phi\}$

EM on Hidden Markov Models



EM ON HIDDEN MARKOV MODELS

Complete log-likelihood function

$Q(\theta|\theta^{old}) = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \theta^{old}) \ln p(\mathbf{X}, \mathbf{Z}|\theta)$

If we define the marginals of $\mathbf{z}_n$ as

$p(\mathbf{z}_n|\mathbf{X}, \theta^{old}) = \gamma(\mathbf{z}_n)$

$p(\mathbf{z}_{n-1}, \mathbf{z}_n|\mathbf{X}, \theta^{old}) = \xi(\mathbf{z}_{n-1}, \mathbf{z}_n)$

Expectation of a binary random variable : the probability that it takes the value 1;

$\gamma(z_{nk}) = E[z_{nk}] = \sum_{\mathbf{z}} \gamma(\mathbf{z}) z_{nk}$

$\xi(z_{n-1j}, z_{nk}) = E[z_{n-1j} z_{nk}] = \sum_{\mathbf{z}} \gamma(\mathbf{z}) z_{n-1j} z_{nk}$

EM ON HIDDEN MARKOV MODELS

The E-step yields

$$Q(\theta|\theta^{old}) = \sum_{k=1}^{K} \gamma(z_{1k}) \ln \pi_k + \sum_{n=2}^{N} \sum_{j=1}^{K} \sum_{k=1}^{K} \xi(z_{n-1j}, z_{nk}) \ln A_{jk} +$$

$$\sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \ln p(\mathbf{x}_n|\phi_k)$$

EM ON HIDDEN MARKOV MODELS

The M-step maximizes $Q(\theta|\theta^{old})$ with respect to θ subject to $\sum\limits_{i=1}^{S} \pi_i = 1$, $\pi_i \geq 0$ and $\sum\limits_{i=1}^{S} a_{ik} = 1$, $a_{ij} \geq 0$ to yield

$$\pi_k = \frac{\gamma(z_{1k})}{\sum\limits_{j=1}^{K} \gamma(z_{1j})}$$

$$a_{jk} = \frac{\sum\limits_{n=2}^{N} \xi(z_{n-1j}, z_{nk})}{\sum\limits_{l=1}^{K} \sum\limits_{n=2}^{N} \xi(z_{n-1j}, z_{nl})}$$

If we choose $p(\mathbf{x}|\phi_k) = \mathcal{N}(\mathbf{x}|\mu_k, \Sigma_k)$

$$\mu_k = \frac{\sum\limits_{n=1}^{N} \gamma(z_{nk})\mathbf{x}_n}{\sum\limits_{n=1}^{N} \gamma(z_{nk})} \text{ and } \Sigma_k = \frac{\sum\limits_{n=1}^{N} \gamma(z_{nk})(\mathbf{x}_n - \mu_k)(\mathbf{x}_n - \mu_k)^T}{\sum\limits_{n=1}^{N} \gamma(z_{nk})}$$