

Formation et Analyse d'Images

James L. Crowley

ENSIMAG 3

2008/2009

Séance 12

23 janvier 2009

PCA, Mélange de Gaussiennes et L'Algorithme EM (Expectation-Maximization)

| | |
|---|----|
| La Reconnaissance de Visage..... | 1 |
| L'Analyse en Composantes Principales..... | 3 |
| Exemple : | 7 |
| Reconstruction..... | 9 |
| Reconnaissance de visages avec PCA..... | 10 |
| Mélange de Gaussiens. | 11 |
| Ebauche de l'Algorithme EM..... | 13 |
| L'Algorithme EM (Expectation-Maximization)..... | 14 |
| L'étape E..... | 15 |
| L'étape M..... | 16 |

Sources Bibliographiques :

"Neural Networks for Pattern Recognition", C. M. Bishop, Oxford Univ. Press, 1995.

"Pattern Recognition and Scene Analysis", R. E. Duda and P. E. Hart, Wiley, 1973.

La Reconnaissance de Visage

Soit M_k images de visages de K personnes $\{W_{mk}\}$, où les images sont normalisées en position et taille. Une telle normalisation peut-être fait, par exemple, en utilisant les premiers et deuxième moments de région de couleur de peau.

Soit $\{W_m\}$ est l'ensemble de tous les M images $\{W_m\} = \bigcup_{k=1}^K \{W_{mk}\}$, $M = \sum_{k=1}^K M_k$

Notre problème :

Ayant une nouvelle image (normalisé) W , notre problème est de déterminer lequel des K individus il correspond. Si possible, nous voudrions avoir, aussi, un facteur de confiance dans cette évaluation.

k : estimer que W est une image de la personne k .

Pour faire cette estimation, nous allons utiliser la règle de Bayes :

$$p(k | X) = \frac{P(X | k) p(k)}{P(X)}$$

$$k = \arg\max_k \{p(k | X)\} = \arg\max_k \left\{ \frac{P(X | k) p(k)}{P(X)} \right\}$$

avec CF_k comme facteur de confiance :

$$CF_k = p(k | X)$$

Mais nous avons deux problèmes :

1) Quel vecteur de caractéristiques X d'utiliser ?

Les composantes principales (i,j)

$$X = \langle W(i,j), (i,j) \rangle = \sum_{i,j} W(i,j) \cdot (i,j)$$

2) Comment représenter $p(k | X)$ et $P(X)$?

Les mélanges de Gaussiens.

$$P(X) = \sum_{n=1}^N \mathcal{N}(X; \mu_n, C_n)$$

et

$$P(X | k) = \sum_{n=1}^N \mathcal{N}(X; \mu_{kn}, C_{kn})$$

L'Analyse en Composantes Principales

L'analyse en composant principales est une méthodes de déterminer un sous-espace "optimale" pour la reconstruction. Il est souvent utilisé pour la discrimination dans la cas où le vecteur X est composé d'une grande nombre de caractéristiques.

Voici son application pour la reconnaissance d'images de visages.

Soit une ensemble de M imagettes, toutes classes confondus, $W_m(i, j)$, composé de N pixels :

$W_m(i, j)$ pour $m \in \{1, M\}$, tel que $i \in [1, I], j \in [1, J], I \times J = N$

On les exprime sous forme de vecteur :

Ensemble d'imagettes $W_m(n) = W_m(i, j)$ ou $n = j \cdot I + i$.

On cherche une base orthogonale $(n) = \{ d(n) \}$ $d = 1, 2, \dots, D$ pour représenter les $X(n)$. telles que $D \ll M$.

Imagette moyenne $\mu(n) = \frac{1}{M} \sum_{m=1}^M W_m(n)$

Imagette moyenne zéro $V_m(n) = W_m(n) - \mu(n)$

Base orthogonal $(n) = \{ d(n) \}$ $d = 1, 2, \dots, D$
avec $D \ll N$.

Vecteur $= \langle V(n), (n) \rangle$

Image reconstruite : $\hat{X}(n) = \mu(n) + \sum_{n=1}^D v_n(n)$

Image de résidu : $R(n) = X(n) - \hat{X}(n)$

Energie de résidu : $r^2 = \sum_{n=1}^D R^2(n)$

La covariance, **C**, est composée de $N \times N = N^2$ termes

$$\mathbf{C} = E\{ \mathbf{V} \mathbf{V}^T \} = E\{ [\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T \}$$

(pour une image de taille $D= 2^N$, il y a 2^{2N} pixels et 2^{4N} termes dans la covariance.)

$$c_{ij}^2 = \frac{1}{M} \sum_{m=1}^M (V_m(i) \cdot V_m(j))$$

On peut faire une matrix **V** composé de **M** colonnes $\{V_m\}$

$$\mathbf{V} = \begin{matrix} V_{11} & V_{12} & \dots & V_{1M} \\ V_{21} & V_{22} & \dots & V_{2M} \\ \dots & \dots & & \dots \\ V_{N1} & V_{N2} & \dots & V_{NM} \end{matrix}$$

$$\mathbf{C}_x = \mathbf{V} \mathbf{V}^T = \begin{matrix} \bullet & \bullet & \bullet & \bullet & & & \\ \bullet & \bullet & \bullet & \bullet & & & \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \\ \bullet & \bullet & \bullet & \bullet & & & \\ \bullet & \bullet & \bullet & \bullet & & & \end{matrix} = \dots$$

V est N par M . Chaque colonne de **V** est une image. Chaque coefficient, c_{ij}^2 , est une covariance de la pixel i et j pour l'ensemble de **M** images

Pour une image de 32×32 , le matrice $\mathbf{C} = \mathbf{V} \mathbf{V}^T$ est de taille 1024×1024 . Il y a un coefficient par pair de pixels. Chaque terme est la covariance d'une paire de pixels.

$$\mathbf{C} = \mathbf{V} \mathbf{V}^T \text{ est de taille } N \times N$$

On cherche une ensemble orthogonales $\{ \mathbf{d}(n) \mid n = 1, 2, \dots, N \}$ tels que : $\mathbf{V}^T \mathbf{V} = \mathbf{c}$

\mathbf{c} est une matrice diagonale des valeurs principales de **C**.

Chaque colonne de V est un vecteur directeur $d(n)$. Les colonnes d sont orthogonales. Nous allons choisir une sous ensemble D des N colonnes de V . Une telle matrice de rotation est fournie par une procédure d'analyse en composants principales.

$$(d(n), c) = \text{PCA}(C).$$

Pour une image de 32×32 , le matrice $C = V V^T$ est de taille $2^{10} \times 2^{10} = 2^{20}$.
 Pour une image de 512×512 , le matrice $C = V V^T$ est de taille $2^{18} \times 2^{18} = 2^{36}$.

Heureusement, il y a une astuce pour éviter la matrice de covariance de $N \times N$ coefficients. Le rank de $V V^T$ est M , ou M est le nombre d'images. $M \ll N$.

Noter que $B = V^T V$ est de taille M^2 , M est le nombre d'images.
 Chaque coefficient est une produit de deux images !

$$V^T V = B = \begin{matrix} \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \end{matrix} = \begin{matrix} \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \end{matrix} = \begin{matrix} \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \\ \bullet & \bullet & \bullet & \bullet \end{matrix}$$

Les coefficients sont les covariances des images (et non pas les pixels!).

$$b_{ij}^2 = \sum_{n=1}^N V_{i(n)} V_{j(n)}$$

Pour $M < 512$ on peut facilement calculer une matrice R de rotation tels que chaque colonne est un vecteur directeur orthogonal.

Soit R les composantes principales de $\tilde{X}^T \tilde{X}$

$$R^T V^T V R = b$$

On multiplie les deux cotés par R :

$$R R^T V^T V R = R b$$

On note que $R^T R = I$.

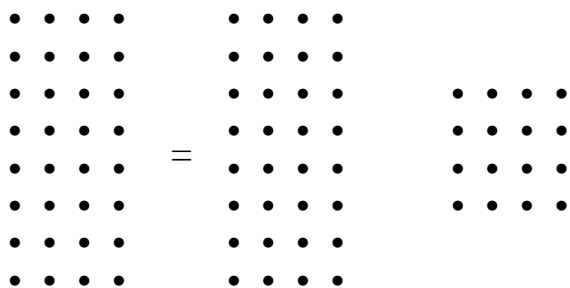
donc :

$$V^T V R = R \quad b$$

Maintenant on multiplie par V

$$\begin{aligned} V V^T V R &= V R \quad b \\ &= (V V^T)(V R) = (V R) \quad b \\ &= (V V^T) \quad = \quad c \end{aligned}$$

Donc, par inspection : $= VR.$
 $= V \quad R$



Raison : Les mêmes images ont été utilisé pour V et pour R .

Les premiers M vecteurs propres de $V^T V$ sont aussi les premiers M vecteurs propres de $V V^T$, le Rank de V est le même que R , et b sont les premiers M valeurs propre de c et Donc, les vecteurs propres : $d(n) = V R$ triés par b

$$= VR$$

Chaque colonne est un vecteur dans une base orthogonale.

$$x^{(n)} = \sum_{m=1}^M V_{m(n)} R(m,n)$$

$x^{(n)}$ fournit une base "ortho-normal" pour $X_m(n)$.

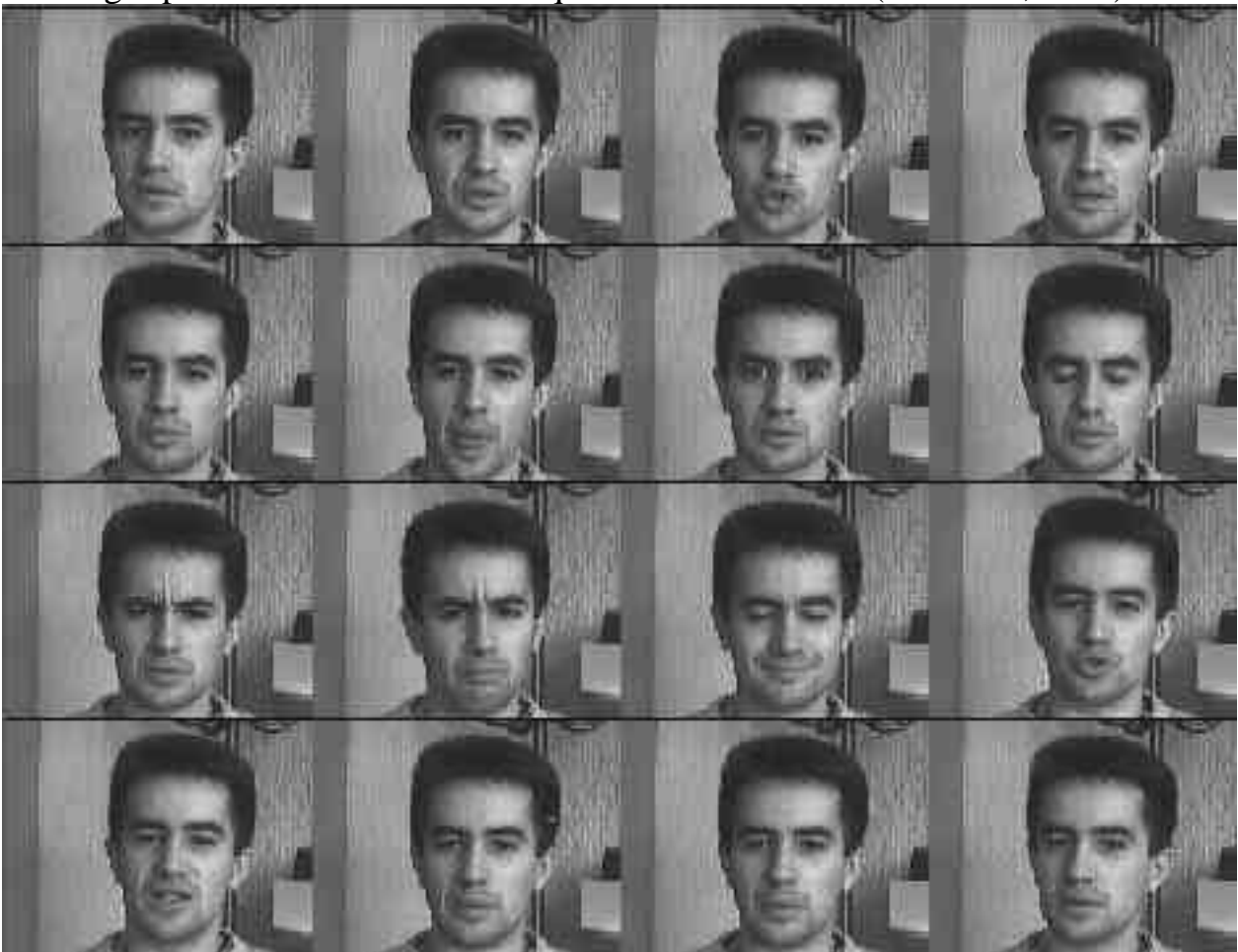
Une image (normalisée) peut être exprimé par

$$d = \langle X(n), d(n) \rangle = \sum_{d=1}^M X(n) d(n)$$

Les valeurs d sont un "code" qui représente $X(n)$ pour la reconnaissance ou la transmission.

Exemple :

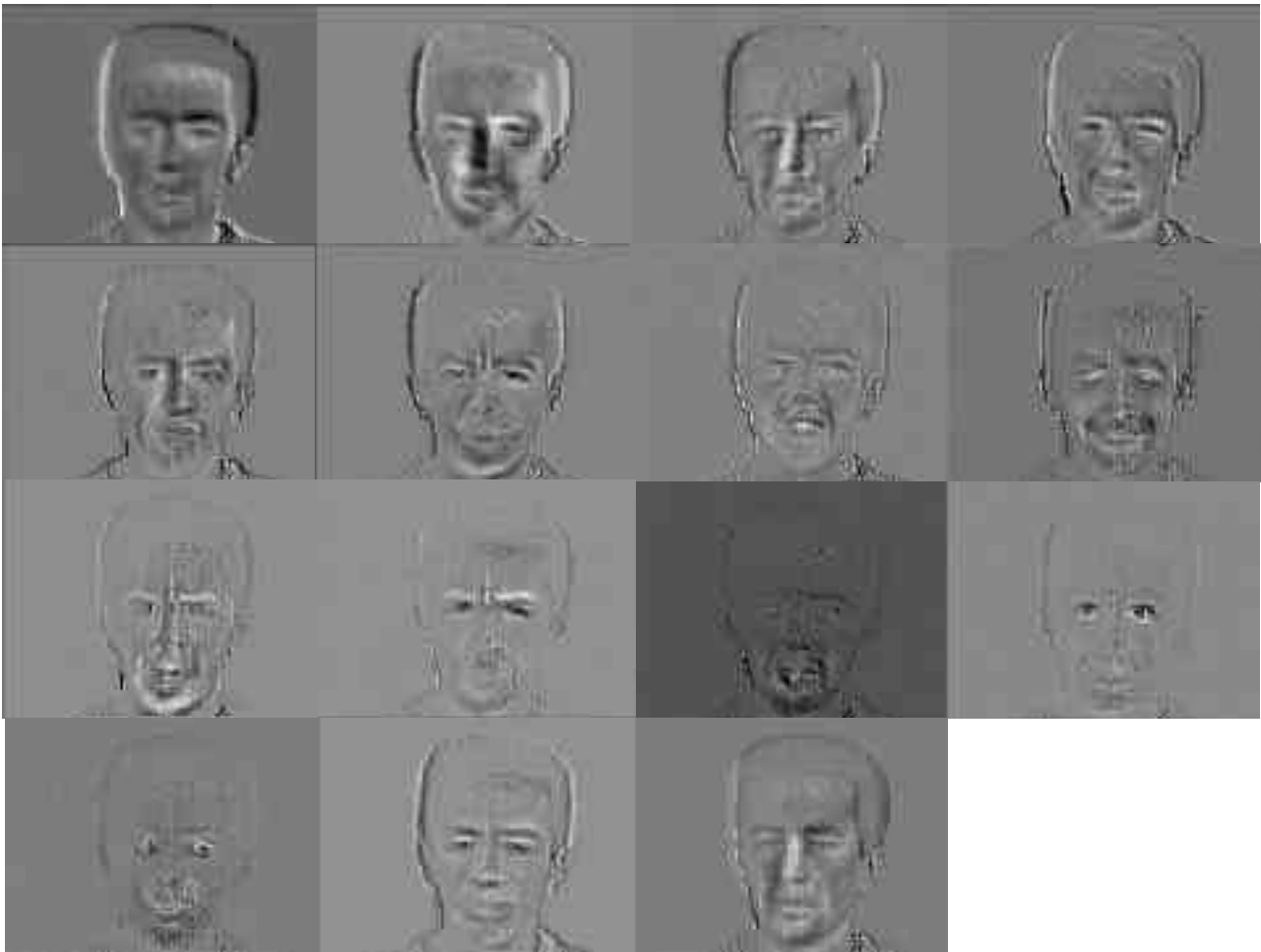
16 images pris au hasard dans une séquence de 2 minutes. (F. Bérard, 1995).



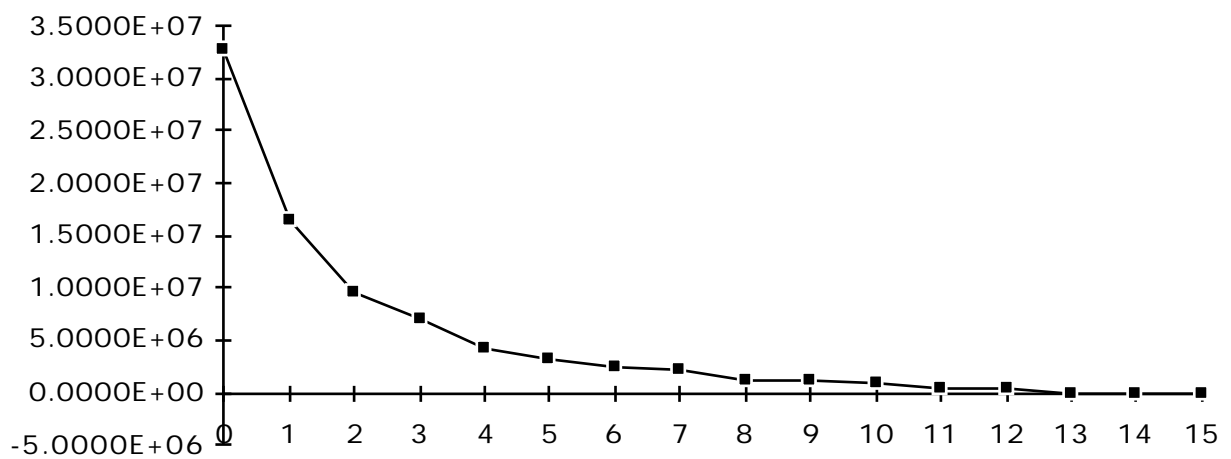
Average Image



Composants Principales :



Eigen Values.



Reconstruction

Image Reconstructed image (120 bytes) Error Image.



Reconstruction (120 bytes)



Image Error

Reconnaissance de visages avec PCA

Dans l'espace PCA, pour chaque individu, on fait un ensemble de M_k images.

Ensemble d'images de la classe k : $W_{mk}(n) = W_{mk}(i, j)$ ou $n = j*32 + i$.

Image moyenne de la classe T_k : $\mu_k(n) = \frac{1}{M} \sum_{m=1}^M W_{mk}(n)$

Image moyenne zéro $V_{mk}(n) = X_{mk}(n) - \mu_k(n)$

Projection sur : $X_{mk} = \langle V_{mk}, \rangle$

La covariance, C_k , est composée de $D_y \times D_y$ termes

$$C_k = E\{ V_{mk} V_{mk}^T \}$$

(pour une image de taille 2^n , il y a 2^{2n} pixels et 2^{4n} termes dans la covariance.)

$$ij^2 = \frac{1}{M} \sum_{m=1}^M (V_{mk}(i) V_{mk}(j))$$

ou bien :

$$ij^2 = \frac{1}{M} \sum_{m=1}^M (W_{mk}(i) - \mu_k(i))(W_{mk}(j) - \mu_k(j))$$

Mélange de Gaussiens.

La reconnaissance est fait par

$$p(k | X) = \frac{p(X | k) p(k)}{p(X)}$$

Si les événements sont issus d'une composition de "N" phénomènes, la densité $p(X)$ prendra la forme d'une composition de lois Normales. Pour les visages, il peut s'agit des variations dans l'angle de l'éclairage ou le fait de porter les lunettes.

Dans un tel cas, on peut approximer par une somme pondérée de densités Normales.

$$p(x) = \sum_{n=1}^N \alpha_n \mathcal{N}(x; \mu_n, \sigma_n)$$

Un tel somme est connu par la terme "mélange de Gaussiens" pour chaque Gaussien il faut estimer trois paramètres :

$$\alpha_n = (\alpha_n, \mu_n, \sigma_n)$$

En total il y a $3 \cdot N$ paramètres à estimer.

$$= (\alpha_1, \mu_1, \sigma_1, \alpha_2, \mu_2, \sigma_2, \dots, \alpha_N, \mu_N, \sigma_N)$$

Si toutes les μ et σ était fixe (et égale), on pouvait calculer les α_n directement.

Mais parce qu'ils sont libres il faut les estimer par un processus itérative.

Un tel processus est composé de deux étapes.

Ces étapes fournissent une estimation des variables cachées.

Pour un mélange de Gaussiens, les variables cachées sont les "sources" pour des événements.

On suppose chaque événement est issu d'un des N sources.

Nous allons construire une table de probabilités. $h(m, n)$

$h(m, n) = \Pr\{\text{l'événement } E_m \text{ est issu de la source } N\}$

les probabilités, $h(m, n)$, nous donnera les facteurs de Mélange, π_n , ainsi que μ_n , σ_n .
L'algorithme d'estimation s'appelle "Expectation-Maximization" ou "EM".

Ebauche de l'Algorithme EM

Soit une ensemble ("training set") de M observations $\mathbf{T} = \{X_m\}$.

Fait une premier estimation des paramètres $^{(0)}$ et puis on alterne "E" et "M".

E: Faire une estimation des valeurs manquantes, $h(m, n)$, pour les événements.

$h(m, n)^{(i)} = p(h_m | X_1, X_2, \dots, X_M, \quad ^{(i)})$ pour chaque terme "n".

$$h(m, n)^{(i)} = \frac{n^{(i)} \mathcal{N}(X_m; \mu_n^{(i)}, \sigma_n^{(i)})}{\sum_{j=1}^N j^{(i)} \mathcal{N}(X_m; \mu_j^{(i)}, \sigma_j^{(i)})}$$

M: Recalculer $^{(i+1)}$ avec $p(h_m | X_1, X_2, \dots, X_M, \quad ^{(i)})$

$$S_n^{(i+1)} = \sum_{m=1}^M h(m, n)^{(i)}$$

$$n^{(i+1)} = \frac{1}{M} S_n^{(i+1)} = \frac{1}{M} \sum_{m=1}^M h(m, n)^{(i)}$$

$$\mu_n^{(i+1)} = \frac{1}{S_n^{(i+1)}} \sum_{m=1}^M h(m, n)^{(i)} X_m$$

$$\sigma_n^{(i+1)} = \frac{1}{S_n^{(i+1)}} \sum_{m=1}^M h(m, n)^{(i)} (X_m - \mu_n^{(i+1)})^2$$

Pour la dérivation l'algorithme EM, il faut introduire le concept de "likelihood" (vraisemblance).

L'Algorithme EM (Expectation-Maximization)

L'algorithme EM s'applique à l'estimation de données cachés.

Il est utilisé notamment pour l'estimation des Modèles de Markov Cachées (HMM's) et pour l'estimation des Mélanges de Gaussiens.

Pour un mélange de Gaussiens, les variables cachées sont les sources des événements.

On suppose que chaque événement est produit par un des N sources.

Pour chaque événement E_m On définit la variable "cachée" h_m

$h_m = n$ si l'événement E_m (avec caractéristique X_m) est issu de la source N.

Nous cherchons à estimer

$$p(x) = \prod_{n=1}^N \mathcal{N}(x; \mu_n, \sigma_n) \quad \text{Tels que} \quad \sum_{n=1}^N \pi_n = 1.$$

Il faut estimer N vecteurs : $\theta_n = (\pi_n, \mu_n, \sigma_n)$

En total il y a $3 \cdot N$ paramètres à estimer.

$$\theta = (\pi_1, \mu_1, \sigma_1, \pi_2, \mu_2, \sigma_2, \dots, \pi_N, \mu_N, \sigma_N)$$

$$l(\theta | \{X_m\}) = \text{Log} (L(\theta | X_1, X_2, \dots, X_M)) = \text{Log} \left\{ \prod_{m=1}^M p(X_m | \theta) \right\}$$

$$= \sum_{m=1}^M \text{Log} \{ p(X_m | \theta) \}$$

$$= \sum_{m=1}^M \text{Log} \left(\sum_{n=1}^N \pi_n p(X_m | \theta_n) \right)$$

Il faut $\hat{\theta}_S = \max_{\theta} \{ l(\theta | \{X_m\}) \}$

Il n'y pas de solution analytique, parce qu'il n'y est pas une dérivé pour la logarithme de la somme.

Soit $\mathbf{T} = \{X_m\}$ $h(m,n)$ le donnée Complète.

On vas maximiser une mesure de qualité $Q(\theta^{(i)})$ définit par une esperence conditionnel

$$Q(\theta^{(i)}) = E \{ l(\theta^{(i)} | \mathbf{T}) \} = \text{Log} \left\{ \prod_{m=1}^M p(X_m | \theta^{(i)}) \right\}$$

L'etape E

Pour chaque événement E_m avec caractéristique X_m ,

On suppose qu'il manque l'information: h_m

$h_m = n$, la source de l'événement E_m .

On ne connaît pas h_m , mais on peut estimer les probabilités $P(h_m = n)$.
par une table $h(m, n)$.

Pour chaque X_m , et son source caché h_m

$$p(h_m, X_m | \cdot) = p(h_m | X_m, \cdot) p(X_m | \cdot)$$

donc

$$p(h_m | X_m, \cdot) = \frac{p(h_m, X_m | \cdot)}{p(X_m | \cdot)}$$

où

$$p(h_m = n, X_m | \cdot) = \mathcal{N}(X_m; \mu_n, \sigma_n)$$

$$p(X_m | \cdot) = \sum_{n=1}^N \mathcal{N}(X_m; \mu_n, \sigma_n)$$

Donc

$$p(h_m = n | X_1, X_2, \dots, X_M, j) = \frac{\mathcal{N}(X_m; \mu_n, \sigma_n)}{\sum_{j=1}^N \mathcal{N}(X_m; \mu_j, \sigma_j)}$$

Donc pour chaque itération (i) le premier étape E est :

$$h(m, n)^{(i)} = \frac{\mathcal{N}(X_m; \mu_n^{(i)}, \sigma_n^{(i)})}{\sum_{j=1}^N \mathcal{N}(X_m; \mu_j^{(i)}, \sigma_j^{(i)})}$$

L'etape M

Pour le deuxième étape, M, nous allons maximiser le "likelihood" $\text{Log}\{p(\mathbf{X}_m | \cdot)\}$
 On définit la "Expected Complete Data Log Likelihood" pour $\mathbf{T} = \{\mathbf{X}_m\} \cup h(m,n)$

$$Q = Q(\cdot^{(i)}) - Q(\cdot^{(i-1)})$$

Pour chaque cycle dans l'itération nous allons chercher :

$$\cdot^{(i+1)} = \text{argmax}\{Q(\cdot | \cdot^{(i)})\}$$

Ce maximum est donné par :

$$S_n^{(i+1)} = \sum_{m=1}^M p(h_{m=n} | \mathbf{X}_m, \cdot^{(i)}) = \sum_{m=1}^M h(m, n)$$

$$n^{(i+1)} = \frac{1}{M} \sum_{m=1}^M P(h_{m=n} | \mathbf{X}_m, \cdot^{(i)}) = \frac{1}{M} S_n^{(i+1)}$$

$$\mu_n^{(i+1)} = \frac{1}{S_n^{(i)}} \sum_{m=1}^M p(h_{m=n} | \mathbf{X}_m, \cdot^{(i)}) \mathbf{X}_m = \frac{1}{S_n^{(i+1)}} \sum_{m=1}^M h(m, n) \mathbf{X}_m$$

$$\begin{aligned} \sigma_n^{(i+1)} &= \frac{1}{S_n^{(i)}} \sum_{m=1}^M p(h_{m=n} | \mathbf{X}_m, \cdot^{(i)}) (\mathbf{X}_m - \mu_n^{(i+1)})^2 \\ &= \frac{1}{S_n^{(i+1)}} \sum_{m=1}^M h(m, n) (\mathbf{X}_m - \mu_n^{(i+1)})^2 \end{aligned}$$

Avec notre table de probabilités $h(m, n)$

$$h(m, n) = P(h_m=n \mid X_m, \quad (i))$$

E (Expectation) :

$$h(m, n)^{(i)} := \frac{n^{(i)} \mathcal{N}(X_m; \mu_n^{(i)}, \sigma_n^{(i)})}{\sum_{j=1}^N j^{(i)} \mathcal{N}(X_m; \mu_j^{(i)}, \sigma_j^{(i)})}$$

M: (Maximisation)

$$S_n^{(i+1)} := \sum_{m=1}^M h(m, n)^{(i)}$$

$$n^{(i+1)} := \frac{1}{M} S_n^{(i+1)}$$

$$\mu_n^{(i+1)} := \frac{1}{S_n^{(i+1)}} \sum_{m=1}^M h(m, n)^{(i)} X_m$$

$$\sigma_n^{(i+1)} := \frac{1}{S_n^{(i+1)}} \sum_{m=1}^M h(m, n)^{(i)} (X_m - \mu_n^{(i+1)})^2$$

Dans le cas Multivariate ($D > 1$) la covariance C est composée de σ_{jk}^2 :

$$\sigma_{jkn}^{(i+1)} := \frac{1}{S_n^{(i+1)}} \sum_{m=1}^M h(m, n)^{(i)} (X_{jm} - \mu_{jn}^{(i+1)})(X_{km} - \mu_{kn}^{(i+1)})$$