Modélisation des Locuteurs Les Modèles de Mélanges Gaussiennes GMMS

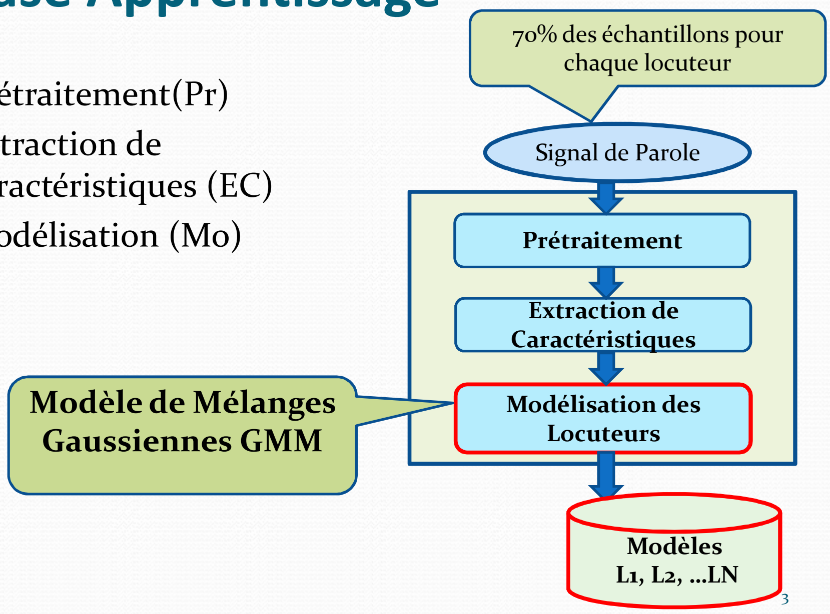
Par Djellali hayet
Master RSI 2020
UBMA Université
Algérie

Sommaire

- Principe des Mélange de Gaussiennes
- Modélisation des GMMs pour la IAL
- L'algorithme Estimation Maximisation EM
- Exemple pratique
- TP IAL par GMMs

RAL
Phase Apprentissage

- Prétraitement(Pr)
- Extraction de caractéristiques (EC)
- Modélisation (Mo)



RAL La modélisation

- **Dépendant du texte**
  - Les modèles de Markov Caches sont mieux adaptés en dépentant du texte
  - Tp avec outil HTK
- **Indépendant du texte**
  - Les Modèles GMMs
  - Les Modèles de Mélanges Gaussiens

Probabilité Conditionnelle

- P(B/A)= P(B) * P (A / B) / P(A)

C’est quoi une Gaussienne?

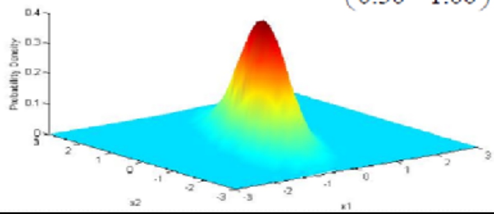
What is a Gaussian?

For ***d*** **dimensions**, the Gaussian distribution of a vector $x = (x^1, x^2, \ldots, x^d)^T$ is defined by:

$$N(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{d/2}\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

where μ is the mean and Σ is the covariance matrix of the Gaussian.

Example: $\mu = (0,0)^T$ $\Sigma = \begin{pmatrix} 0.25 & 0.30 \\ 0.30 & 1.00 \end{pmatrix}$



5

Code de la fonction de densité probabilité

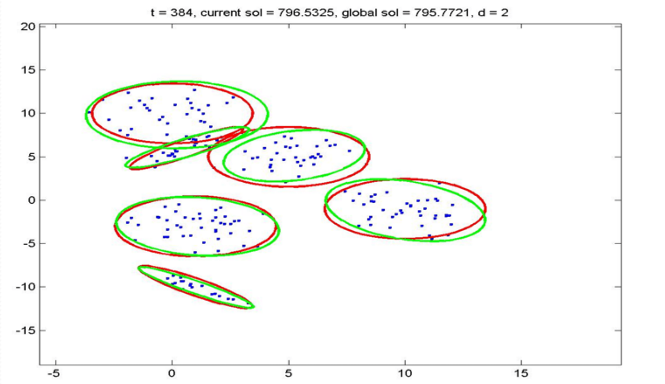
- function b = **gaussian_pdf**(x, mu, sigma);
- // x: vecteur de caracteristiques (MFCC ou autre)
- // mu : moyenne de x,   sigma : matrice de covariance
- d = length(mu);
- c = (2*pi)^(d/2);
- c = c*(det(sigma)^(1/2));
- for i=1:size(x,2)
- **b(i,:)** = **exp**((-1/2)*((x(:,i)-mu)'*inv(sigma)*(x(:,i)-mu)));
- end
- b = b./c;
- return

Le Problème

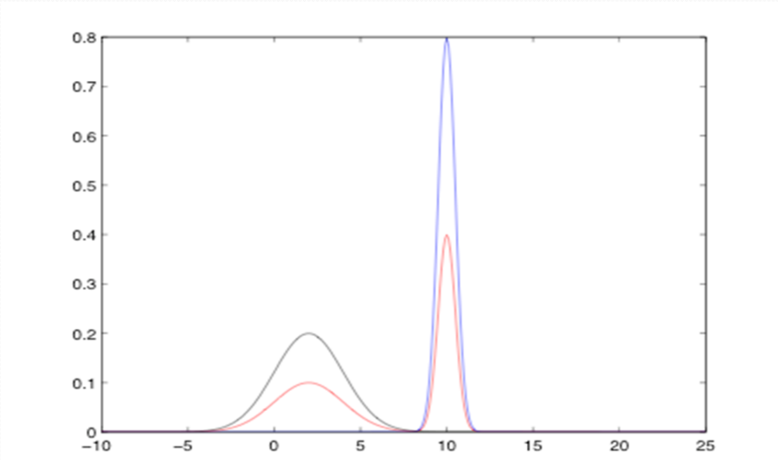
- Vous avez des Données d'une population de taille N.
- Vous Cherchez a connaitre les parametres de chaque population.
- Vous ne savez rien sur cette population à priori.
- Vous pensez juste qu'elle obeit à une gaussienne.

Les Modèles de Mélange Gaussiennes

- Au lieu d'identifier les clusters par 'les plus proches' centroids.
- Distribuez un ensemble de *k* Gaussiannes au données.
- Maximum de vraissemblance (Maximum Likelihood) pour un modèle de mélanges.



Exemple de deux GMM f0 et f1



$f_0(x) = N(x; 2, 2)$ $\pi = [.5 \quad .5]^T$ $f_1(x) = N(x; 10, .5)$

Modèle de Mélanges

- Un modèle de mélange est la somme pondérée de plusieurs PDF(fonction de densité de probabilité) ou les poids sont déterminés par la distribution π .

$$p(x) = \pi_0 f_0(x) + \pi_1 f_1(x) + \pi_2 f_2(x) + \ldots + \pi_k f_k(x)$$

$$\text{where } \sum_{i=0}^{k} \pi_i = 1 \qquad \boxed{p(x) = \sum_{i=0}^{k} \pi_i f_i(x)}$$

Gaussian Mixture Models

- GMM: La Somme pondérée de K+1 gaussiennes ou les poids sont déterminés par la distribution.

π

$$p(x) = \pi_0 N(x|\mu_0, \Sigma_0) + \pi_1 N(x|\mu_1, \Sigma_1) + \ldots + \pi_k N(x|\mu_k, \Sigma_k)$$

$$\text{where } \sum_{i=0}^{k} \pi_i = 1$$

$$\boxed{p(x) = \sum_{i=0}^{k} \pi_i N(x|\mu_k, \Sigma_k)}$$

Mélange de Gaussiennes GMMs

- Un mélange de Gaussiennes est une somme pondérée de M densités gaussiennes.
- Soit un locuteur S et un vecteur acoustique x de dimension D, le mélange de gaussiennes est défini comme suit :
- $p(x|\lambda_s) = \sum_{m-1}^{M} \pi_m^s \, b_m^s(x)$
- Où les $b_m^s(x)$ représentent des densités gaussiennes, paramétrées par un vecteur de moyenne μ_m^s et une matrice de covariance $\sum_m^s$:

Gaussienne

$$b_m^s(x) = \frac{1}{(2\pi)^{D/2} . |\sum_m^s|^{1/2}} \, exp\left[-\frac{1}{2}(x - \mu_m^s)'(\sum_m^s)^{-1}(x - \mu_m^s)\right]$$

$$\mathcal{N}(x|\mu_m, \Sigma_m) = \frac{1}{\sqrt{(2\pi)^D|\Sigma_m|}} \, exp\left(-\frac{1}{2}(x-\mu_m)^T \, \Sigma_m^{-1} \, (x-\mu_m)\right)$$

Et les π_m^s représentent les poids du mélange, avec $\sum_{m-1}^{M} \pi_m^s = 1$.

Modèle du Locuteur

- Un locuteur est donc modélisé par un ensemble de paramètres noté λ_S :

- $\lambda_S = \{\pi_m^S, \mu_m^S, \sum_m^S\}_{m-1,\ldots,M}$
- π_m^s représentent les poids du mélange,

avec $\sum_{m-1}^{M} \pi_m^s = 1.$

μ_m^s : Vecteur Moyenne

$\sum_m^s$: une matrice de covariance :

Mélange de Gaussiennes GMMs

- Mélange Gaussiens: $p(x|\lambda_s) = \sum_{m-1}^{M} \pi_m^s \, b_m^s(x)$ (1.9)
- Les Densités Gaussiennes: $b_m^s(x)$
- Le vecteur de moyenne μ_m^s
- Une matrice de covariance $\sum_m^s$:
- $b_m^s(x) = \frac{1}{(2\pi)^{D/2} . |\sum_m^s|^{1/2}} \times$

$$exp\left[-\frac{1}{2}(x-\mu_m^s)'(\textstyle\sum_m^s)^{-1}(x-\mu_m^s)\right]$$

- Les poids du mélange π_m^s, avec $\sum_{m-1}^{M} \pi_m^s = 1.$
- Un locuteur est donc modélisé par un ensemble de paramètres noté λ_S :
- $\lambda_S = \{\pi_m^s, \mu_m^s, \sum_m^s\}_{m-1,\ldots,M}$

Mélange de Gaussiennes GMMs

- Ce modèle de mélanges gaussiennes GMM peut prendre plusieurs formes. Concernant les matrices de covariance:
- On peut affecter une matrice de covariance à chaque gaussienne, ou bien utiliser une matrice de covariance globale, commune à toutes les gaussiennes.
- Elles peuvent être pleines ou diagonales (en raison de la faible corrélation des coefficients melcepstraux.
- On considèrera généralement les matrices de covariance diagonales) [MAM, 2003].

Modélisation des GMMs pour la RAL

- Depuis leur introduction par Reynolds en 1992, les modèles GMM sont devenus l'état de l'art en modélisation de locuteur en mode indépendant du texte.
- Cette approche consiste à modéliser un Locuteur par un Mélange de Gaussiennes qui représente une somme pondérée de M gaussiennes multi-dimensionnelles.
- Chaque gaussienne gi est caractérisée par son **poids pi**, un vecteur **moyen µi** de dimension d et une matrice de **covariance ∑i** de dimension d x d [REY, 1992] [REY, 1994b].

Modélisation des GMMs pour la RAL

- Cette approche a donné, depuis plus de 10 ans, de bonnes performances pour les systèmes de reconnaissance du locuteur en mode indépendant du texte [MAA, 2006] et les meilleurs résultats ont été atteints avec l'extraction des coefficients MFCC dans l'étape d'extraction des paramètres,
- Il a été démontré que les GMM sont d'un haut niveau de fiabilité dans des systèmes pareils [WIL, 2001].
- Les travaux de D. A. Reynolds (1995) constituent l'état de l'art en la matière [MAM, 2003].

Modélisation des GMMs pour la RAL

- L'utilisation d'un modèle GMM se justifie en faisant appel à l'interprétation des classes du mélange : il est certain que les vecteurs de paramètres vont se répartir différemment selon les caractéristiques du son de parole considéré (**son voisé / non voisé**, ou plus finement en **fonction du phonème**).
- Chaque composante va modéliser des ensembles sous-jacents de classes acoustiques, chaque classe représentant des évènements acoustiques (voyelles, nasales, ...).
- L'allure spectrale de la ième classe pourra être représentée par la moyenne et la matrice de covariance de la ième composante.
- Ces classes caractérisent l'espace acoustique propre à chaque locuteur.
- La raison poussant à utiliser les GMM est qu'à l'aide d'une combinaison linéaire de Gaussiennes, on peut représenter une large gamme de distributions [MAM, 2003].

Mélange de Gaussiennes GMMs

- L'apprentissage d'un modèle GMM consiste en l'estimation de l'ensemble des paramétres λ = {μm, Σm, wm} , M=1, en utilisant un ensemble de données d'apprentissage X ={x1, x2, ..., xT} (xt ∈ RD).
- Cet apprentissage fait souvent appel `a la technique d'estimation par maximum de vraisemblance (Maximum Likelihood Estimation) MLE [Fisher, 1925] ;
- on utilise souvent l'algorithme Espérance Maximisation (Expectation-maximisation) EM [Dempster et al., 1977], [Bishop, 2006].

Apprentissage par l'Algorithme Espérance Maximisation (Expectation-maximisation) EM

- **répéter**
- **Step1**:

  Calculer la probabilité a posteriori
- **Step2**:
- Recalculer : Poids W; Moyenne mu

  Matrice de Covariance Sigma
- Jusqu'à convergence.

Estimation Maximisation EM

- Calculez les probabilités a posteriori

$$P(m|x_t) = \frac{w_m \mathcal{N}(x_t|\mu_m, \Sigma_m)}{\sum_{g=1}^{M} w_g \mathcal{N}(x_t|\mu_g, \Sigma_g)}.$$

Estimation Maximisation EM

- L'´etape de maximisation (M), ou on réestime les paramétres du modèle afin de maximiser la vraisemblance :

$$w_m = \frac{1}{T}\sum_{t=1}^{T} P(m|x_t),$$

$$\mu_m = \frac{\sum_{t=1}^{T}\left(P(m|x_t)x_t\right)}{\sum_{t=1}^{T} P(m|x_t)},$$

$$\Sigma_m = \frac{\sum_{t=1}^{T}\left(P(m|x_t)(x_t-\mu_m)(x_t-\mu_m)^T\right)}{\sum_{t=1}^{T} P(m|x_t)}.$$

Estimation Maximisation EM

- **function** [mixtures,p_i] = **gmmem8j(X,mixtures);**
- N = size(X,2) ; %longueur signal 10000
- dim=size(X,1);%13 coef ; M = length(mixtures);
- p_i = ones(1, M)/M; ; minEigenvalue = 2;
- % disp('ubm female 8 speakers 8 mixtures....');
- **for o = 1:20; For i=1:M**
  gm = mixtures(i);
  Q(:, i) = gaussian8j_pdf(X, gm.mu, gm.sigma)* p_i(i);
- %**p(xd|i) * p(i)**
- End
- su = sum(Q, 2); Q = Q./repmat(su, [1 M]);
- **for i=1:M**
- gm = mixtures(i); **sum_p_i_x = sum(Q(:,i));**
- **p_i(i) = (1/size(Q,1))*sum_p_i_x**; **gm.mu = X*Q(:,i)./sum_p_i_x;**
- **sigma = zeros(dim);**

Suite Estimation-Maximization

- **for j=1:N**
- V = X(:,j) - gm.mu; h=Q(j,i)*(V'*V);
- **sigma = sigma +h ;**
- **end**
- oldsigma = gm.sigma; **gm.sigma = sigma/sum_p_i_x**;
- [P, L] = eig(gm.sigma);
- if any(diag(L)<minEigenvalue)
- gm.sigma = oldsigma;
- end
- **mixtures(i) = gm**;
- end
- disp('o= '); disp(num2str(o)); plot(Q(:,1:M),'r.');
- end

- Merci de Votre Attention