Gaussian Mixture Model Classifiers
Applications to MIR

Bertrand SCHERRER

February 7, 2007

Outline

1. Key Concepts
   - GMM : an “Unsupervised Classifier"
   - Why “Gaussian Mixture" ?
2. Practical example of GMMs applied to MIR
   - Context
   - GMM Training
   - Classification test
3. Other Applications
   - Musical Instrument Identification in Polyphonic Music
   - Extraction of melodic lines from audio recordings
4. Conclusion

Outline

1. Key Concepts
   - GMM : an “Unsupervised Classifier"
   - Why “Gaussian Mixture" ?
2. Practical example of GMMs applied to MIR
   - Context
   - GMM Training
   - Classification test
3. Other Applications
   - Musical Instrument Identification in Polyphonic Music
   - Extraction of melodic lines from audio recordings
4. Conclusion

GMM : an “Unsupervised Classifier"

Unsupervised Classifier

The training samples of the classifier are not labelled to show their category membership [Duda 73].

Advantages

- Less time consuming when applied to a large set of data.
- Ability to track (slow) time-evolving patterns.

GMM : an “Unsupervised Classifier"

Unsupervised Classifier

The training samples of the classifier are not labelled to show their category membership [Duda 73].

Advantages

- Less time consuming when applied to a large set of data.
- Ability to track (slow) time-evolving patterns.

Why “Gaussian Mixture" ?

In GMM classifier, for a given class, the probability density function of the observation vector is modelled as :

$$p(\mathbf{x}|C_i) = \sum_{k=1}^{K} P(k|C_i).G_k(\boldsymbol{\mu_k}, \Sigma_k) \tag{1}$$

where :

- $\mathbf{x}$ is a d-component feature vector.
- $\boldsymbol{\mu_k}$'s are the d-component mean vectors of Gaussian G_k.
- Σ_k's are the d-by-d covariance matrices of Gaussian G_k.
- $P(k|C_i)$ is the *a priori* probability of Gaussian G_k for instrument class C_i.

Why “Gaussian Mixture" ?

In GMM classifier, for a given class, the probability density function of the observation vector is modelled as :

$$p(\mathbf{x}|C_i) = \sum_{k=1}^{K} P(k|C_i).G_k(\boldsymbol{\mu_k}, \Sigma_k) \quad (1)$$

where :

- $\mathbf{x}$ is a d-component feature vector.
- $\boldsymbol{\mu_k}$'s are the d-component mean vectors of Gaussian G_k.
- Σ_k's are the d-by-d covariance matrices of Gaussian G_k.
- $P(k|C_i)$ is the *a priori* probability of Gaussian G_k for instrument class C_i.

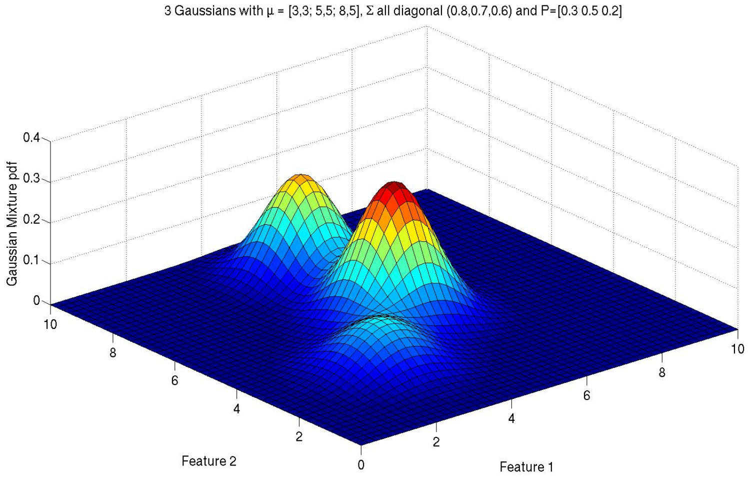
Why “Gaussian Mixture" ?

In GMM classifier, for a given class, the probability density function of the observation vector is modelled as :

$$p(\mathbf{x}|C_i) = \sum_{k=1}^{K} P(k|C_i).G_k(\boldsymbol{\mu_k}, \Sigma_k) \tag{1}$$

where :

- $\mathbf{x}$ is a d-component feature vector.
- $\boldsymbol{\mu_k}$'s are the d-component mean vectors of Gaussian G_k.
- Σ_k's are the d-by-d covariance matrices of Gaussian G_k.
- $P(k|C_i)$ is the *a priori* probability of Gaussian G_k for instrument class C_i.


3 Gaussians with μ = [3,3; 5,5; 8,5], Σ all diagonal (0.8,0.7,0.6) and P=[0.3 0.5 0.2]
Gaussian Mixture pdf
0.4
0.3
0.2
0.1
0
Feature 2
10
8
6
4
2
0
Feature 1
2
4
6
8
10

Outline

- GMMs are used in many different fields.
- Look at one clear example of GMM classification applied to MIR: [Marques 99]

- GMMs are used in many different fields.
- Look at one clear example of GMM classification applied to MIR: [Marques 99]

Context

Objectives

Instrument identification in monophonic music:

- 8 different “classes": bagpipe, clarinet, flute, harpsichord, organ, piano, trombone and violin .
- on very short recordings (0.2s).

What is to be classified ?

A set **X** of *m* **unlabelled** observations (cepstral, mel-cepstral and LPC coefficients) :

- $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2 \ldots \mathbf{x}_m]$.

Context

Objectives

Instrument identification in monophonic music:

- 8 different “classes": bagpipe, clarinet, flute, harpsichord, organ, piano, trombone and violin .
- on very short recordings (0.2s).

What is to be classified ?

A set **X** of *m* **unlabelled** observations (cepstral, mel-cepstral and LPC coefficients) :

- $\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2} \dots \mathbf{x_m}]$.

Context

Assuming that the observations are i.i.d., the **likelihood** that the entire set of observations $\mathbf{X}$ has been produced by a violin (C_0 for example) is :

$$p\left(\mathbf{X}|C_0\right) = \prod_{t=1}^{m} p(\mathbf{x_t}|C_0) \tag{2}$$

and each $p(\mathbf{x_t}|C_0)$ is modelled as a mixture of K multivariate gaussians.

Objectives of the GMM training

At this stage, one tries to estimate, for all the classes of instruments, the parameters of the GMM:

- $\theta_{ik} = [P(k|C_i), \mu_{k,i}, \Sigma_{k,i}]$ for $k = 1 \ldots K$.

How to estimate the GMM parameters : MLE ?

- Ideal way would be to use the **Maximum Likelihood Estimation** (a.k.a. MLE).
- MLE theoretically consists in finding $\theta = [\theta_{i1}, \theta_{i2} \dots \theta_{iK}]$, maximizing $p(\mathbf{X}|C_i)$.
- In the case where all the parameters are unknown, MLE becomes **very** complex ... and unreliable

$\Rightarrow$ need for an **alternate method.**

How to estimate the GMM parameters : MLE ?

- Ideal way would be to use the **Maximum Likelihood Estimation** (a.k.a. MLE).
- MLE theoretically consists in finding $\boldsymbol{\theta} = [\boldsymbol{\theta}_{i1}, \boldsymbol{\theta}_{i2} \ldots \boldsymbol{\theta}_{iK}]$, maximizing $p(\mathbf{X}|C_i)$.
- In the case where all the parameters are unknown, MLE becomes **very** complex ... and unreliable

$\Rightarrow$ need for an **alternate method.**

How to estimate the GMM parameters : MLE ?

- Ideal way would be to use the **Maximum Likelihood Estimation** (a.k.a. MLE).
- MLE theoretically consists in finding $\boldsymbol{\theta} = [\boldsymbol{\theta}_{i1}, \boldsymbol{\theta}_{i2} \ldots \boldsymbol{\theta}_{iK}]$, maximizing $p(\mathbf{X}|C_i)$.
- In the case where all the parameters are unknown, MLE becomes **very** complex ... and unreliable

$\Rightarrow$ need for an **alternate method.**

How to estimate the GMM parameters : MLE ?

- Ideal way would be to use the **Maximum Likelihood Estimation** (a.k.a. MLE).
- MLE theoretically consists in finding $\boldsymbol{\theta} = [\boldsymbol{\theta}_{i1}, \boldsymbol{\theta}_{i2} \ldots \boldsymbol{\theta}_{iK}]$, maximizing $p(\mathbf{X}|C_i)$.
- In the case where all the parameters are unknown, MLE becomes **very** complex ... and unreliable

$\Rightarrow$ need for an **alternate method.**

How to estimate the GMM parameters : EM

- **Expectation Maximisation** algorithm [Dempster 77] is an iterative solution very often used for MLE.

Initial Steps

1. Compute closed-form expressions of the parameters of the GMM corresponding to a local extremum of the likelihood $p(\mathbf{X}|\theta(t))$.
2. Make a first guess on the values of $\theta(t)$.

$\Rightarrow$ Let's iterate !

How to estimate the GMM parameters : EM

- **Expectation Maximisation** algorithm [Dempster 77] is an iterative solution very often used for MLE.

Initial Steps

1. Compute closed-form expressions of the parameters of the GMM corresponding to a local extremum of the likelihood $p(\mathbf{X}|\theta(t))$.
2. Make a first guess on the values of $\theta(t)$.

$\Rightarrow$ Let's iterate !

How to estimate the GMM parameters : EM

- **Expectation Maximisation** algorithm [Dempster 77] is an iterative solution very often used for MLE.

Initial Steps

1. Compute closed-form expressions of the parameters of the GMM corresponding to a local extremum of the likelihood $p(\mathbf{X}|\theta(t))$.
2. Make a first guess on the values of $\theta(t)$.

$\Rightarrow$ Let's iterate !

Classification Test

- A feature vector $\mathbf{x_t}$ is said to belong to an instrument class i if it maximizes $p(C_i|\mathbf{x_t}) = p(\mathbf{x_t}|C_i).p(C_i)$.
- If classes can occur with the same probability, since we know θ for all classes, $\mathbf{x_t}$ belongs to the class for which $p(\mathbf{x_t}|C_i)$ is maximum.

In [Marques 99]:

- **Features :** mel cepstral feature vectors (16-element vectors).
- **Order of the GMM:** 2.

$\Rightarrow$ overall error rate of 37%.

Outline

1. Key Concepts
   - GMM : an “Unsupervised Classifier"
   - Why “Gaussian Mixture" ?
2. Practical example of GMMs applied to MIR
   - Context
   - GMM Training
   - Classification test
3. Other Applications
   - Musical Instrument Identification in Polyphonic Music
   - Extraction of melodic lines from audio recordings
4. Conclusion

Musical Instrument Identification in Polyphonic Music

- In [Eggink 03], try to recognize 2 instruments playing at the same time.
- Approach based on estimation of multiple fundamental frequencies to sort features fed into the GMM.

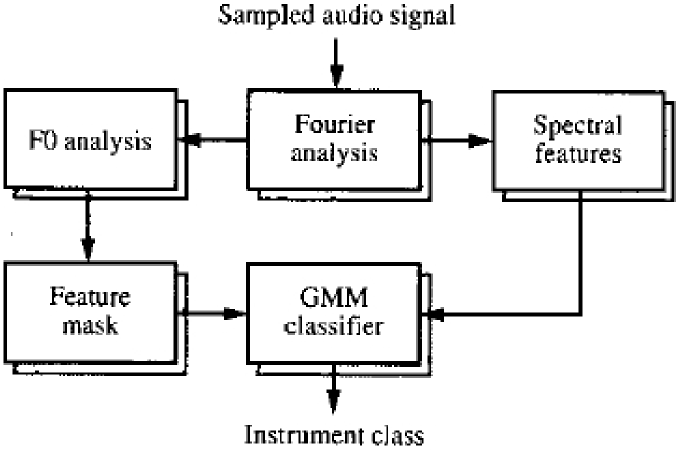


Figure 1 [Eggink 03]

- **Features :** cepstral coefficients clearly belonging to different tones.
- **Order of the GMM:** 120.
- **Training Material:**
  - monophonic recordings of tones or melodies produced by 5 instruments (flute, oboe, clarinet, violin, cello).
  - 1min recordings from different sources.
  - silence was removed.
- **Training Method:** EM algo initialized by k-nearest neighbour.

$\Rightarrow$ Monophonic OK, short duet OK

- **Features :** cepstral coefficients clearly belonging to different tones.
- **Order of the GMM:** 120.
- **Training Material:**
  - monophonic recordings of tones or melodies produced by 5 instruments (flute, oboe, clarinet, violin, cello).
  - 1min recordings from different sources.
  - silence was removed.
- **Training Method:** EM algo initialized by k-nearest neighbour.

$\Rightarrow$ Monophonic OK, short duet OK

Extracting Melodic Lines from Audio Recordings

In [Marolt 04]:

- Use SMS $\Rightarrow$ identify partials.
- Include masking effect.
- Predominant pitch estimation $\Rightarrow$ melodic fragments.
- Use of GMM $\Rightarrow$ group melodic fragments together to form melodic lines

- **Features :** 5 including loudness, pitch stability ...
- **Order of the GMM:**
- **Training Material:**
  - entire song (Respect by Ottis Redding, as sung by Aretha Franklin) that had been annotated by hand.
- **Training Method:** EM with equivalence constraints.

$\Rightarrow$ Identification of the main melody works.

- **Features :** 5 including loudness, pitch stability ...
- **Order of the GMM:**
- **Training Material:**
  - entire song (Respect by Ottis Redding, as sung by Aretha Franklin) that had been annotated by hand.
- **Training Method:** EM with equivalence constraints.

$\Rightarrow$ Identification of the main melody works.

Outline

- GMM widely used in MIR, with many variations : EM algo or not, initialization of the training ...
- Classifiers = complex and fascinating subject.

Questions ?