David Sears
MUTH 621
10/8/09

Gaussian Mixture Models

In this summary I will discuss a common model for the automatic classification of musical data: the Gaussian Mixture Model (GMM). I will first outline a few of the issues GMMs attempt to resolve. I will then discuss how Gaussian distributions classify the data, including the application of the Expectation-Maximization Algorithm to provide an estimate of maximum likelihood. I will conclude with a discussion of some of the applications of GMMs in music information retrieval.

1 Classifying Musical Data

1.1 The Problem

The extraction of meaningful information from acoustic data represents a potentially intractable problem in MIR research. In order to derive methods for the extraction and classification of musical information (e.g. melody, harmony, instrument classification), researchers have turned to machine learning techniques. The goal of machine learning algorithms is to provide a method for the automatic discovery of regularities or similarities in data [6]. Machine learning methods therefore require the presence of 1) a population of data, 2) the extraction of meaningful features from that population, and 3) a model for the classification of the population based on feature information.

1.2 Classifying According to Statistical Principles: The Probability Density Function

In order to make statistical inferences in classification, or in any parametric test, scientists employ a Gaussian (normal) distribution to estimate the observed data. The Gaussian distribution is a continuous probability distribution that describes data that cluster around a mean. A probability density function (pdf) therefore provides a theoretical estimate of a sample of data; it is thought of as the density according to which a population is distributed.

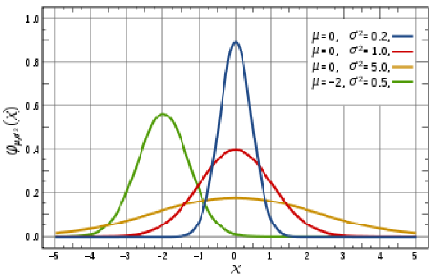


Figure 1, open source, wikipedia.org

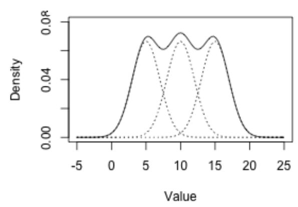
2 Gaussian Mixture Models

2.1 Mixture Models

A Gaussian Mixture can simply be understood as a number of Gaussians introduced into a population of data in order to classify each of the possible sample clusters, each of which could refer to instrument category, melodic line, etc. GMMs are an example of an unsupervised learning algorithm, because none of the sample points are labeled (which is statistically analogous to a pdf).

2.2 Maximum Likelihood Estimation: EM Algorithm

Figure 2, open source, wikipedia.org



Maximum likelihood Estimation (MLE) is a method for fitting a statistical model to the data. It roughly corresponds to least-squares error, shown in the equation below:

Standard Error = $\sqrt{\sum}(x-\mu)^2$

MLE therefore provides a means of determining how well the function fits the observed data. In the case of a GMM, however, the data is clumped together as a mixture of possible Gaussian functions, so the ML estimate cannot be determined in the face of missing or hidden data. The Expectation-Maximization Algorithm provides a solution to this problem. It is an iterative procedure which consists of two processes: the E-step, in which the missing data are estimated given the observed data and the current estimate of the model parameters, and the M-step, in which the likelihood function is maximized under the assumption that the missing data are known (thanks to the E-step). At each iteration the algorithm converges toward the ML estimate [3]. In the E-step, the initial Gaussian parameters (means, covariances) are generated via the k-means method [8].

3 Applications to Music

3.1 Instrument Classification

In [8], the authors use a GMM to classify instruments in sound segments of .2 seconds in length. The authors extracted 3 features: linear prediction features (attempts to approximate the musical instrument system), cepstral features (the first components of the cepstrum correspond to the production model or shape of the spectrum), and mel cepstral features (this feature filters the cepstral features according to how a human perceives the frequency content). In a training phase, the EM algorithm computes the ML estimates for the mixing weights and mean vectors (ie the parameters) of the Gaussian mixture. The authors reported that the mel cepstral feature set gave the best results, with an overall error rate of 37%.

3.2 Melodic Lines

In [7], the author employed a GMM to classify melodic lines from an Aretha Franklin recording of “Respect.” The EM algorithm iteratively estimated the weights of all of the possible pitches at a given segment, and then represented the dominant tone in the observed PDF. She reported that for lead vocals, the GMM classified to an accuracy of .93, while for the bass line it was more accurate (.97).

4 Conclusions

GMMs provide a common method for the classification of data whose class weights are unknown. Considering the complexity of the model, the importance of choosing relevant features cannot be overestimated, as any classifier depends on the features in order to determine similarities or differences in the data set.

References

[1] Balkema, W. 2007. Variable-Size Gaussian Mixture Models for Music Similarity Measures. Proc. International Symposium on Music Information Retrieval (ISMIR) , pp. 491- 494, Vienna, Austria, September 2007.

[2] Bilmes, J. 1998. A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. Technical Report ICSI-TR-97-021, University of Berkeley.

[3] Borman, S. 2004. The Expectation Maximization Algorithm. Retrieved October 3, 2009, from http://www.seanborman.com/publications/EM_algorithm.pdf. Version: 2004. [n.p.].

[4] Dellaert, F. 2002. The Expectation Maximization Algorithm. Technical Report GIT-GVU-02-20, College of Computing, Georgia Institute of Technology, February 2002.

[5] Dempster, A. P. Laird, N.M. & Rubin, D. B. 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society. Series B.. 19(1): 1-38.

[6] Eck, D. Eck's Presentations on Machine Learning & Classification. Retrieved on October 3, 2009, from http://www.iro.umontreal.ca/%7Epift6080/H09/documents/intro_ml.pdf. [n.p.].

[7] Marolt, M. 2004. Gaussian Mixture Models for Extraction of Melodic Lines from Audio Recordings. Proc. 5th Int. Conf. on Music Information Retrieval (ISMIR'04), pp. 80-83.

[8] Marques, J. & Moreno, P. 1999. A Study of Musical Instrument Classification Using Gaussian Mixture Models and Support Vector Machines. Tech. Rep., Cambridge Research Laboratory, Cambridge, Mass, USA, June 1999.

[9] Moon, T. 1996. The Expectation-Maximization Algorithm. IEEE Trans. Signal Process. 44, 1, 47â€“59.