Final Report: A Machine Learning Approach to Classifying Music Genre

Jon Whelpley

November 19, 2020

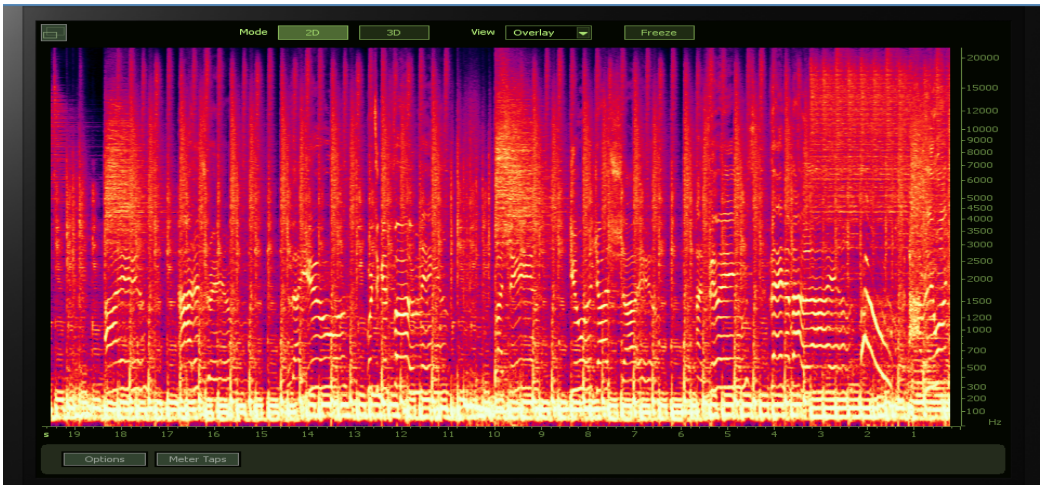
Contents

Abstract

Music is often classified into subjective categories called genres. People typically classify songs based on qualities such as tempo, key signature, lyrical context, timbre, etc. This purpose of this project is to investigate machine learning classification methods to accurately predict what genre a song belongs to given features associated with the song. The Free Music Archive provides a data set with 8000 30 second samples of songs that fall into 8 preset genres that are later partitioned into testing and training sets. They also provide meta data containing a list of features for each song. This information is used to generate the models for classification.

1 Introduction

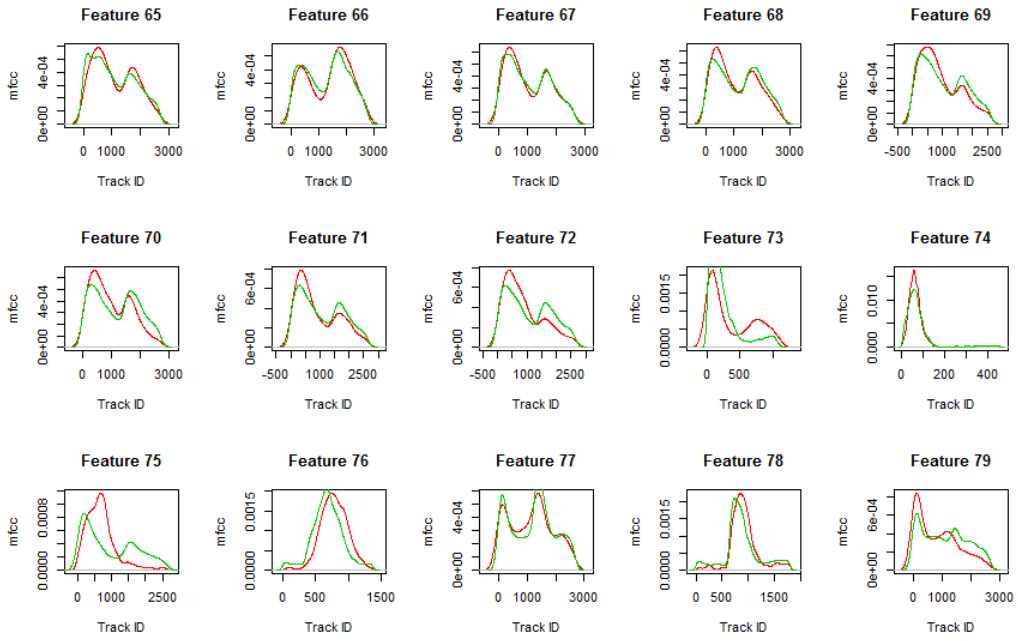
Statistical Learning is the set of tools for modeling and understanding complex data sets. It is a continuously growing field of research that has applications in any field requiring reliable predictions. Because of the significance and abundance of big data it has become necessary to consider such applications. Machine learning has grown out of statistical learning due to the immediate applications in computer science. The scope of this research is to construct a model using a statistical learning approach to classify music genre.



2 The Data Set

The data set used came from the Free Music Archive's small data set of 8000 songs pre-classified into 8 genres: Electronic, Experimental, Folk, Hip-Hop, Instrumental, International, Pop, and Rock. The features for each song were extracted via the LibROSA package in Python. Such features include metrics from chromagrams, mel frequency cepstral coefficients (MFCCs), and Root mean square energy (RMSE). In total, we used 162 of the features made available from the FMA metadata.

We separate the data set into a set of training data, which for this model contains 5000 of the 8000 songs. These were chosen randomly in R. We used 2989 of the remaining 3000 as test data to determine model accuracy. We also re-sampled the training and testing data to help gauge how sensitivity of the model.



This is a sample of the distribution of the features for the FMA data set. One curve represents the distribution of a particular feature associated with songs in the Rock genre while the other correlates to the distribution of that feature with songs that do not fall under Rock.

3 Preliminaries

Suppose we have a set of K classes to classify objects into. We want to create decision boundaries in order to decide whether or not a given object belongs to a specific class. In the context of classification of music into genres, it is necessary to define a way to "separate" the genres. We can do this with the following approaches.

3.1 Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a method of classification that estimates the probability that an object x belongs to a given class based on its predictors. In other words for each object that is fed into the model we must find $Pr(Y = k|X = x)$. We can do this with a few assumptions of normality and Bayes' Theorem for Classification.

Bayes' Theorem states

$$Pr(Y = k|X = x) = \frac{\pi_k f_k(x)}{\sum_{l=1}^{K} \pi_l f_l(x)}$$

where

- π_i is the probability that x is from the i^{th} class, and
- $f_i(x) = Pr(X = x|Y = i)$

If we assume we have a single predictor per object we may be able to make a selection for what $f_k(x)$ might be. In particular if we assume the predictors of the object have a normal distribution we can say

$$f_k(x) = \frac{1}{\sqrt{2\pi}\sigma_k} exp(-\frac{1}{2\sigma_k^2}(x - \mu_k)^2)$$

and furthermore we can say

$$Pr(Y = k|X = x) = \frac{\pi_k \frac{1}{\sqrt{2\pi}\sigma_k} exp(-\frac{1}{2\sigma_k^2}(x - \mu_k)^2)}{\sum_{l=1}^{K} \pi_l \frac{1}{\sqrt{2\pi}\sigma_l} exp(-\frac{1}{2\sigma_l^2}(x - \mu_l)^2)}$$

.

If we let the number of predictors p greater than 2, x now becomes a vector and we can define a new f given by the multivariate normal distribution.

$$f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} exp(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu))$$

where

- Σ is the co-variance of x
- μ is the expectation of x
- $|\Sigma|$ denotes the determinant of Σ

Using Bayes' Theorem and Algebraic Manipulation we find that an object will be classified into the class where $\delta_k(x)$ is the highest where

$$\delta_k(x) = x^T\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^T\Sigma^{-1}\mu_k + log(\pi_k)$$

3.2 Support Vector Machines

Support Vector Machines (SVMs) are used to generate a separating hyper-plane that divides the two classes given margins. Finding the separating hyper-plane reduces to the optimization problem

$$\max_{\beta_0,...,\beta_p,\epsilon_0,...,\epsilon_p,M} M$$

$$\text{s.t. } \sum_{j=1}^{p} \beta_j^2 = 1$$

$$y_i(\beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip}) \geq M(1-\epsilon_i)$$

$$\epsilon_i \geq 0, \sum_{i=1}^{n} \epsilon_i \leq C$$

- Where C is a non negative "tuning" parameter
- M is the width of the margin around the hyper-plane
- Each ϵ_i is a slack variable that allows for misclassification of training data.

Classification then reduced down to finding out which side of the hyperplane a new object falls into. In other words if we have an object x^* we only need to see if $f(x^*) = \beta_0 + \beta_1 x_1^* + ... + \beta_p x_p^*$ is positive or negative.

This particular method of using a support vector classifier generates a linear decision boundary, however; sometimes the boundaries are non linear and require a more nuanced approach. It turns out that solutions to the optimization problem are related to the inner products observations. Hence one can find the linear support vector classifier to be

$$f(x) = \beta_0 + \sum_{i \in S} \alpha_i \langle x, x_i \rangle$$

.

More generally, we can have alternative classification methods by constructing a kernel $K(x, x_i)$ to replace the inner product in the above solution for $f(x)$.

Some examples of Kernels include:

Linear:

$$K(x_i, x_{i'}) = \langle x_i, x_{i'} \rangle$$

Radial:

$$K(x_i, x_{i'}) = exp(-\gamma \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2)$$

Polynomial:

$$K(x_i, x_{i'}) = (1 + \sum_{j=1}^{p} x_{ij} x_{i'j})^d$$

3.3 Principal Component Analysis

This is a method for dimension reduction of features. I.e. if we notice linear dependence of features we may be able to limit the number of feature vectors to a smaller set of new vectors to reduce memory constraints. We would ideally like to find a small set of vectors that still contain most of the information of the original feature space. The new normalized vectors, also known as the principal components are of the form:

$$Z_i = \phi_{i1}X_1 + \phi_{i2}X_2 + ... + \phi_{ip}X_p$$

Where p is the number of predictors and the ϕ_{ji}'s are called loadings. Since the vectors are normalized, we have the following property:

$$\sum_{j=1}^{p} \phi_{ji}^2 = 1$$

So to find a principal component one must maximize over the parameters $\phi_{1i}, \phi_{2i}, ..., \phi_{pi}$

$$\frac{1}{n}\sum_{k=1}^{n}\left(\sum_{j=1}^{p}\phi_{ji}x_{kj}\right)^2$$

4 Approach and Methodology

Our initial goal was to find an appropriate classifier that yielded satisfactory results and then to extract more features from the audio to improve model accuracy. We accomplished the first goal through surveying a few classification methods and evaluating the results. Initially, the results were dis satisfactory so we felt it was appropriate the simplify the problem to a binary classification problem since the classification methods were developed for binary classification. We surveyed the same classification methods, tried to find which features had the most influence on classification accuracy, and adjusted some parameters of the classifiers to eventually reach our model.

5 Results

5.1 Initial Results for Multi-classification

All Genre Classifier With Support Vector Machines and Principal Component Analysis (Linear Kernel)

Genre:	Elec.	Exp.	Folk	Hip-Hop	Instr.	Internat.	Pop	Rock
Classified as Electronic	58	66	41	63	18	84	49	43
Classified as Experimental	51	22	24	45	18	32	68	100
Classified as Folk	45	83	92	38	88	35	43	56
Classified as Hip-Hop	48	49	35	16	36	39	37	22
Classified as Instrumental	20	46	62	33	96	33	45	25
Classified as International	65	47	33	43	44	32	58	84
Classified as Pop	20	24	41	51	24	63	29	27
Classified as Rock	62	53	55	80	37	63	37	13

All Genre Classifier With Linear Discriminant Analysis and Principal Component Analysis

Genre:	Elec.	Exp.	Folk	Hip-Hop	Instr.	Internat.	Pop	Rock
Classified as Electronic	58	60	31	76	12	66	44	39
Classified as Experimental	57	29	28	53	28	52	79	121
Classified as Folk	45	74	62	63	50	36	71	73
Classified as Hip-Hop	62	48	25	21	26	40	37	24
Classified as Instrumental	29	59	87	26	138	34	38	22
Classified as International	49	35	18	27	24	17	29	52
Classified as Pop	22	25	54	35	36	77	26	20
Classified as Rock	47	60	78	68	47	59	42	19

The first table yields 11.98 percent accuracy and the second yields 12.38 percent accuracy. With these kinds of numbers our classification methods would just be equivalent to simply guessing the genre randomly. At this point we decided to tackle a simpler problem in hopes of being able to scale it to the multi-classification problem.

5.2 Binary Classification Results

When we reduced the problem to the binary case we selected one genre, namely Rock, to compare to the rest of the genres which were categorized as "Not Rock." We initially used the same methods from before and quickly found they were just as ineffective. We then attempted to select individual features that might have the most influence on classification accuracy. We found that the features that influence classification the most were the MFCCs extracted from each track. We also used a function in the Random Forest Package in R to find which features were most significant in classifying via Random Forest (which was experimentally verified to be consistent with the variables that yielded the most accurate classification results via SVM).

Binary Classification Results with LDA

Genre:	Not Rock	Rock
Classified as Not Rock	2569	343
Classified as Rock	76	1

Binary Classification Results with SVM (Linear Kernel)

Genre:	Not Rock	Rock
Classified as Not Rock	2504	342
Classified as Rock	141	2

Binary Classification with Random Forest

Genre:	Not Rock	Rock
Classified as Not Rock	2607	257
Classified as Rock	38	87

Binary Classification with Variable Selection and SVM

Genre:	Not Rock	Rock
Classified as Not Rock	2564	202
Classified as Rock	47	176

Eventually, we started experimenting with Kernels and reached a confusion matrix that seemed to be close to what we were initially looking for. More Rock songs were being classified as Rock.

Binary Classification with SVM (Radial Kernel)

Genre:	Not Rock	Rock
Classified as Not Rock	2468	186
Classified as Rock	134	201

5.3 An Adjustment to the Model

After finding a somewhat satisfactory model for the Binary Classification problem we found that scaling the adjustments to the Multi-Classifier proved to yield far better results than what we previously had.

All Genre Classifier with Support Vector Machines and All Features (Radial Kernel)

Genre:	Elec.	Exp.	Folk	Hip-Hop	Instr.	Internat.	Pop	Rock
Classified as Electronic	233	32	4	49	26	21	23	8
Classified as Experimental	13	156	17	15	30	13	32	21
Classified as Folk	6	38	251	7	38	19	39	25
Classified as Hip-Hop	54	23	1	236	7	32	26	11
Classified as Instrumental	26	63	33	8	193	13	22	21
Classified as International	17	15	21	22	5	235	23	7
Classified as Pop	49	49	37	42	33	33	147	55
Classified as Rock	9	20	8	11	15	8	22	221

6 Conclusion

The results of our multi- classifier are by no means novel. Others have produced models with greater accuracy and more sophisticated classification methods and feature extraction techniques. However, looking at the confusion matrix for the most accurate model, we find that even though the classification rate was around 56 percent the errors almost align very well with the qualitative boundaries that separate genres. In other words, the errors occur mostly where two particular genres are similar. For example, most of the electronic tracks that were misclassified went to Hip-Hop, Instrumental, and Pop. It's entirely possible that those particular songs could also be considered to be under the "misclassfied" genres.

7 Future Work

In the future it might be beneficial to expand the data set to the large data set provided by the free music archive. It would also be worth considering other features that one could extract from the audio files to improve the model.

References

[1] Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. An Introduction into Statistical Learning With Applications in R. *Springer* (2013), 138-150, 344-356.

[2] Riccardo Petitti, Paolo Annesi, Roberto Basili, Raffaele Gitto, Alessandro Moschitti. Audio Feature Engineering for Automatic Music Genre Classification. (2007).

[3] FMA Data set Link: https://github.com/mdeff/fma