## Introduction

Music classification has been a prominent topic in the field of machine learning in the past decade or so due to the commercializing potentials it has. Personalizing music recommendations based on the users’ history is an increasingly important task for entertainment companies to keep their subscribers loyal to their product [1].

The goal of this project can be broken down into two tasks: perform PCA on the features dataset to observe for any underlying patterns within the dataset that could potentially help differentiate between the genres and if we can construct an LDA model that produce well separated classes analogous to the four genres. The result from this project shows that features such as contrast and the second mel-frequency cepstral coefficient were important to identifying the Classical and Historical/Old-Time genres.

## Data and Features Extraction

### Dataset

The dataset of interest for this project was originally a part of the Kaggle’s “Music Genre Classification” challenge hosted by Rob Mulla as part of NVDIA’s machine learning competition[2]. The audio files were taken from the training dataset that contains 20,000 30-seconds royalty free music samples that were pre-labeled from 0 to 18 indicating their musical genre. For visualization purposes, I am only considering audio files that falls within the following five genres: Rock, Hip-Hop, Classical, Old-Time/Historical, and Jazz. These five genres where specifically chosen because I am interested in observing how well Linear Discriminant Analysis can separate the audio files into their appropriate class, given that the dataset contains music from overlapping genres such as Classical and Old-Time/Historical while also working with samples from the opposite ends of the music spectrum such as Rock and Classical. This subsetting procedures reduced the initial 20,000 samples available in the pre-labeled training dataset down to only 5750 audio files.

Table 1: Audio Files Count Per Musical Genre

|  |  |
| --- | --- |
| **Genre** | **Number of Audio Files** |
| 1 - Rock | 3096 |
| 4 - Hip-Hop | 1761 |
| 11 - Classical | 495 |
| 12 – Old-Time/Historical | 408 |

### Feature Extraction

Audio features for each of the 5750 30-seconds samples were extracted from the ogg formatted audio files using the Librosa package in python. Using the rolling-window method discussed in Holt[3], 8 features were extracted for each 5 seconds window frames resulting in a total of 48 temporal dependent features. Due to the dataset containing a mixture of instrumental and vocal tracks, Mel-frequency cepstral coefficients (MFCC) were extracted to maximize the differentiation between genres that are historically vocal heavy versus those that are more instrumental base. As a final preprocessing step, I appregated the features from their vector forms into singular values by taking the expected values of each feature at the different time frames. Aggregating the features allows us to minimize the dimensionality of the initial feature space we are working in while still capturing enough information to build a fingerprint for each audio sample. Table 2 below contains a list of the extracted features used in this project.

Table -2: List of Extracted Audio Features

|  |  |
| --- | --- |
| **Column Index** | **Features** |
| 2 – 7 | Average Spectral Contrast |
| 8 – 13 | Average Harmonic |
| 14 – 19 | Chroma Shift |
| 20 – 25 | Average Root-Mean Square (RMSE) |
| 26 – 31 | Average Spectral Centroid |
| 32 – 37 | Average Spectral Bandwidth |
| 38 – 43 | Average Spectral Rolloff |
| 44 | Average Zero Cross Rate (zcr) |
| 45 - 58 | Average Mel-Frequency Cepstral Coefficients |

**Spectral Contrast:** Thespectral contrast capture information about the differences in energy between the peaks and valley within a specific frequency sub-interval, usually highlighting the differences between the extreme signal values within the sub-spectrum [4]. Spectral contrast plays a crucial role in identifying patterns within songs that relies on contrasting dynamics.

**Harmonic:** Harmonic can be thought of as the pitched sound we perceive from an audio [5], where each pitch is denoted by a frequency (hz) depending on the octave they exist in. We can think of harmonic as the foundation to constructing a cord progression, where each step in a progression is a different pitch either ascending or descending the ladder.

**Chroma Shift:** Chroma Shift provides information on the melodic progression of a song by representing the changes in pitches within a timeframe as a sequence of chroma features [6], better understood as musical notes. This feature is useful for identifying genres that often are written in a particular set of keys or tend to reuse similar melody while being invariant to changes in octave or instrumentation[7].

**Root-Mean Square:** Root-Mean Square measures (rms) the average of a continuous fluctuating energy output within the audio files, this feature is often used to measure the loudness of the audio at a given timeframe [8].

**Spectral Centroid:** The spectral centroid provides information at which frequency the energy of the spectrum centers around [9], it is often in audio processing as the measure of the musical timbre contained in a song [10].

**Spectral Bandwidth:** The spectral bandwidth captures the oscil

oscillation

the differences between the maximum and minimum frequencies about the spectral centroid, the oscillation between the upper and lower frequencies provide information on the frequencies band

features provides information on the characteristics of the frequencies band for an audio file.

**Mel-Frequency Cepstral Coefficients:** Mel-Frequency Cepstral Coefficients (MFCCs), first introduced by Mermelstein in 1976 , commonly used to perform speech and audio recognition[11]. MFCCs are a set of coefficients that provide an overall shape of the spectral envelope of an audio file [12], often time, MFCCs is use in music classification to gather information on the timber and key of a song [13].

## Methodology

### Principal Component Analysis

Principal Component Analysis (PCA) is a linear dimension reduction technique that aims to minimize the sum of squared residual errors between the original data points in the high-dimensional feature space and the projected data in the lower-dimensional space [14]. The objective of PCA is to find a lower-dimensional representation that captures most of the variations found in the data to minimize the loss of information in the projected data. PCA reduces the dimensionality of the data by describing the data using new features called principal components (PCs). These PCs are a set of orthogonal linear combinations of the original explanatory variables, and each of the PCs contain a vector that shows the contribution of the variables to a particular PC [15]. The PCs are outputted in descending order based on the proportion of variance in the original data explained by the PC [15]. The number of PCs is limited to either the number of explanatory variables, p, given that multicollinearity does not exist or the number of observations, n, in the dataset if n < p. Since the first q numbers of PCs capture most of the variations in the dataset, the remaining p – q PCs can be treated as noise and be discarded without losing too much information from the original data[16]. Therefore, we can use the first q PCs to adequately represent the data in a lower-dimensional space .

The following is a step-by-step generalized process of performing principal component analysis on a high dimensional dataset:

Given a data matrix **X** = [x1, x2, x3, …, xn]

1. Normalizing **X:** Each variable is centered by subtracting out its mean and then divided by its standard deviation.
2. Perform SVD on to get a close-rank approximation of :
3. Compute the Principal Components:

* The diagonal entries of are:
* The column of = [] are the eigenvalue corresponding to

### Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is another linear dimension reduction technique that is popular in classification applications. The main objective of LDA is to find a linear projection that maximizes the separations of the classes in the dataset [14]. The technique projects the data onto a lower-dimension that maximizes the ratio between the between-class variance and the within-class variance [17], therefore, we want to simultaneously maximize the between-class variability and minimizes the within-class variability [14]. LDA begin by normalizing the data by centering and scaling it to minimize the influence of any potential outliers then computes the global centroids, , of all the classes in the data. It then computes the between-class variance and the within-class , where μ denote the projected center, is the projected centroid of class j, and are the ith observation of class j. Finally, LDA solves the generalized eigenvalues problem , where is the largest generalized eigenvalue of () that maximizes the the ratio between the between-class variance and the within-class variance . We can use the eigenvalue, , and eigenvector, , to classify new observations . Using Baye’s theorem, we want to compute for the highest conditional probability that a new observation belonging to a certain class Cj, P(Cj|X = ). To find the highest conditional probability we would need to maximize the following conditional probability: , where denotes the conditional density function and denotes the prior probability of class j. The prior probability of class j, , is proportional to the proportion of observations belonging to class j. It turns out, maximizing P(Cj|X = ) is equivalent to maximizing the discriminant function: [18][19]. Using the class posterior calculated from maximizing the discriminant function, we can classify the new observations into the appropriate classes.

## Application and Results

Two sets of the audio extracted features dataset were used for this project. The first dataset contains the extracted audio features for all the samples for the four genres that were pre-labeled in the original training dataset from Kaggle. The second dataset creates a uniformly distributed subset of the first dataset by randomly taking 408 samples from each genre to work with.

### Exploratory Analysis and Normalization of Audio Dataset using PCA

For this project, PCA is used to perform the following two tasks: data-preprocessing and exploratory analysis. PCA is first performed on the original high-dimensional audio features data to denoise and only keep the most important features for later classification procedure using LDA. Before performing PCA on the original data, we first centered and normalized each of the variables. Since the variables in the dataset have differing scaling and magnitude, it is important to normalize the variables to have the same variance to allow all variables to have the same importance rather than letting variables with large variance dominating the first q PCs. In addition, the data was centered to ensure that the PCs explains the variations in the data rather than the overall mean of the data. This pre-processing helps highlight any prominent features that can later when doing genre classifications using LDA.

PCA was also used to give insights into any important features that can help differentiate the music genres from each other. As seen in figure 1 below, features such as the average second MFCC and average contrast are important to describing Classical and Historical/Old-Time music while the rest of the other features are important for Rock and Hip-Hop music, this trend can be seen in both the original and uniformly distributed dataset. In addition, we see that the first principal component is associated with features related to the timbre and loudness of the samples such as average spectral centroids and average contrast. While the second principal component direction seems to be more related to the harmony of the song since most of the high contributing features of the component are the average harmonics.

Interestingly, we see a deviation between the two datasets when it comes to the contribution of the loading factors in the second principal component. In figure 2B, we see that 21 out of the top 30 features exceed the expected contribution percentage for principal component 2 which make them all important contributor for the component. However, when looking at figure 3B, we see that only 7 out of the top 30 features exceed the expected contribution percentage for the second component. This difference in the number of contributing features for the second principal component could indicate that either the size of the data or the distribution of each genre influences the overall variation in the dataset. To support that, we see that the variance explained by the first component for the original dataset, shown in figure 2c, is 47.5%, while the variance explained by the first component for the uniformly distributed dataset, shown in figure 3c, is 55.1%. This difference in the variance explained by the first principal component between the two datasets could potentially leads to difference class separation when performing LDA on the two datasets.

|  |  |
| --- | --- |
| A graph showing different colored dots  Description automatically generated  Figure A: PCA plot for the original music features dataset using the first two principal components | A graph with red and black lines  Description automatically generated  Figure 1B: Variable loadings plot for the first two Principal Components of the dimensionally reduced original dataset |
| A graph showing different colored dots  Description automatically generated  Figure 1C: PCA plot for the uniformly distributed music features dataset using the first two principal components | A graph with red and black lines  Description automatically generated  Figure 1D: Variable loadings plot for the first two Principal Components of the dimensionally reduced uniformly distributed dataset |

Figure 1: PCA plots for both the original and uniformly distributed music features dataset.

|  |  |
| --- | --- |
| A graph of a number of variable  Description automatically generated  Figure 2A: The bar plot shows the relative contribution from top 30 features for principal component 1using the original non-uniformly distributed dataset. Since all the variables larger than the expected average contribution shown by the red dashed line, we consider these features important contributors to the component. | A graph with blue and red lines  Description automatically generated  Figure 2B: The bar plot shows the relative contribution from top 30 features for principal component 2 using the original non-uniformly distributed dataset. For this component we see that 21 out of the 30 top features exceeds the expected average contribution percentage, with zero crossing rate and harmonic features contributing the most. |
| A graph with blue lines and numbers  Description automatically generated  Figure 2C: Scree plot showing the percentage of variance explained by each component in PCA. Approximately 69% of the variation in data can be explained by the first three principal components. | |

Figure 2: Contribution and Scee plots for the first two principal components of the original dataset. The dataset has been scaled and centered before performing principal component analysis on it.

|  |  |
| --- | --- |
| A graph of a number of variable  Description automatically generated  Figure 3A: The bar plot shows the relative contribution from top 30 features for principal component 1 of the uniformly distributed dataset. Since all the variables larger than the expected average contribution shown by the red dashed line, we consider these features important contributors to the component. | A graph with blue and red lines  Description automatically generated  Figure 3B: The bar plot shows the relative contribution from top 30 features for principal component 2 of the uniformly distributed dataset. For the second component we see that only 7 out of the 30 top features exceed the expected average contribution percentage, the average harmonic features contributing the most. |
| A graph with numbers and a line  Description automatically generated  Figure 3C: Scree plot showing the percentage of variance explained by each component in PCA. Approximately 75.5% of the variation in data can be explained by the first three principal components. | |

Figure 3: Contribution and Scee plots for the first two principal components of the uniformly distributed dataset. The dataset has been scaled and centered before performing principal component analysis on it.

### Classification of Audio Features data using LDA

Genre Classification was done on the dimensionally reduced datasets using LDA to observe how well the model could separates the genres given the features that were extracted from the 30-seconds audio samples. When we do a visual comparison of the LDA plots for the original dataset and the uniformly distributed dataset, shown in figure 4 and figure 5 below, we see that the overall global separation between genres is shared between the two classifying models. However, figure 5 shows using the uniformly distributed dataset produced better separations between the genres along with a reduction in the amount of overlapping between the Classical and Historical/Old-Time clusters and the Rock/Hip-Hop cluster. In addition, figure 5 shows that the observations in both the Classical and Historical/Old-Time classes are more concentrated around their perspective class centroid. The differences shown in figures 4 and 5 could indicate that the priors for the classes have a significant influence on the overall separation between the genres. It makes sense for there to be less separation between the classes in figure 4 since the priors for the Rock and Hip-Hop classes are 0.54 and 0.30 respectively while the priors for Classical and Historical/Old-Time are only 0.086 and 0.071. The differences in the magnitude of the priors between the classes could constitute for misclassification since certain samples from one class that somewhat have similar features to the bigger two classes are more likely to be pulled toward the bigger cluster. With the uniformly distributed data, there are equal probability to by classify into any one of the four genres, LDA have an easier time distinguishing between the classes based since the priors no longer have significant influence on classifying the observations. The influence of the priors on classifying the samples is evidence in the accuracy between the two LDA models as observed in table 3 and table 4. In table 4 we can observe that the proportion of misclassifications between the genres are less than the proportion seen in table 5. In addition, the percentage of correctly classified samples for the LDA model using the uniformly distributed data was 82% while the one using the original data was 78%. However, we cannot confidently say that the changed in the priors solely contributed to the increased in classification accuracy between the two LDA models, this increase could be caused by the reduction in the samples sized used to construct the second model.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| A diagram of different colored dots  Description automatically generated  Figure 4: LDA plot for the original dataset with the centroids labeled for each genre | Table 3: Confusion Matrix for Genre Classification using the Original Dataset   |  |  |  |  |  | | --- | --- | --- | --- | --- | |  | 1- Rock | 4 – Hip-Hop | 11 – Classical | 12 - Historical | | 1- Rock | 2631 | 367 | 83 | 14 | | 4 – Hip-Hop | 632 | 1104 | 14 | 2 | | 11 – Classical | 94 | 10 | 386 | 5 | | 12 - Historical | 15 | 0 | 10 | 383 | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| A chart with colorful dots  Description automatically generated  Figure 5: LDA plot for the uniformly distributed dataset with the centroids laveled for each genre | Table 4: Confusion Matrix for Genre Classification using the Uniformly Distributed dataset   |  |  |  |  |  | | --- | --- | --- | --- | --- | |  | 1- Rock | 4 – Hip-Hop | 11 – Classical | 12 - Historical | | 1- Rock | 198 | 84 | 23 | 1 | | 4 – Hip-Hop | 75 | 226 | 5 | 0 | | 11 – Classical | 9 | 4 | 290 | 3 | | 12 - Historical | 5 | 0 | 10 | 291 | |

## Discussion

### 5.1 Conclusion

### 5.2 Limitations

The main limitation of this project is a consequence of the decision to take the average of each feature for each of the 5 seconds sub-intervals rather than working with the full spectrum outputted by the Librosa package. As discussed previously, taking the average of each feature in each sub-intervals allows us to work with a lower-dimensional feature space while also capturing the global trends of the audio features across the 30 seconds’ timeframe. However, Jiang discussed how the averaged spectral features does not capture the relative spectral representation for each sub-interval which he said is crucial for differentiating between the musical genres [20]. As seen in the results, LDA had difficulty in distinguishing Rock and Hip-Hop music from each other since both genres shares too much of the global characteristics to properly differentiate using the averaged dataset.

## Citation

[1] A. C. M. da Silva, M. A. N. Coelho, and R. F. Neto, “A Music Classification model based on metric learning applied to MP3 audio files,” *Expert Syst. Appl.*, vol. 144, p. 113071, 2020, doi: https://doi.org/10.1016/j.eswa.2019.113071.

[2] R. Mulla, “Music Classification.” Kaggle, 2022. [Audio Files]. Available: https://kaggle.com/competitions/kaggle-pog-series-s01e02

[3] G. Holt and Kavindu Kusal, “Music Genre Classification,” Brigham Young University, Course Report. Accessed: May 11, 2024. [Online]. Available: https://www.google.com/url?sa=t&source=web&rct=j&opi=89978449&url=https://cap.stanford.edu/profiles/cwmd%3Ffid%3D301672%26cwmId%3D10838&ved=2ahUKEwiOk7qK8oaGAxX2EzQIHdE9BqQQFnoECBsQAQ&usg=AOvVaw0EFXv0vjfeIEEeVAnwLbVX

[4] Yuu, “【Wave analytics】What is Spectral Contrast?,” Zenn. Accessed: May 12, 2024. [Online]. Available: https://zenn.dev/yuto\_mo/articles/7413ca2ed4eb5f

[5] D. FitzGerald, “Harmonic/Percussive Separation Using Median Filtering,” in *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, Graz, Austria, 2010, pp. 246–253.

[6] “Chroma feature,” *Wikipedia*. Feb. 09, 2024. Accessed: May 12, 2024. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Chroma\_feature&oldid=1205169278

[7] J. Jogy, “How I Understood: What features to consider while training audio files?,” Medium. Accessed: May 12, 2024. [Online]. Available: https://towardsdatascience.com/how-i-understood-what-features-to-consider-while-training-audio-files-eedfb6e9002b

[8] S.-K.-M. arvindpdmn, “Audio Feature Extraction,” Devopedia. Accessed: May 11, 2024. [Online]. Available: https://devopedia.org/audio-feature-extraction

[9] Nagesh Singh Chauhan, “Audio Data Analysis Using Deep Learning with Python (Part 1),” KDnuggets. Accessed: May 12, 2024. [Online]. Available: https://www.kdnuggets.com/audio-data-analysis-using-deep-learning-with-python-part-1

[10] E. Schubert, J. Wolfe, and A. Tarnopolsky, “Spectral centroid and timbre in complex, multiple instrumental textures,” Aug. 2004.

[11] T. Pál and D. T. Várkonyi, “Comparison of Dimensionality Reduction Techniques on Audio Signals.,” in *ITAT*, 2020, pp. 161–168. [Online]. Available: https://ceur-ws.org/Vol-2718/paper04.pdf

[12] E. Deruty, “Intuitive understanding of MFCCs,” Medium. Accessed: May 12, 2024. [Online]. Available: https://medium.com/@derutycsl/intuitive-understanding-of-mfccs-836d36a1f779

[13] T. L. Li and A. B. Chan, “Genre classification and the invariance of MFCC features to key and tempo,” in *International Conference on MultiMedia Modeling*, Springer, 2011, pp. 317–327.

[14] J. P. Cunningham and Z. Ghahramani, “Linear dimensionality reduction: Survey, insights, and generalizations,” *J. Mach. Learn. Res.*, vol. 16, no. 1, pp. 2859–2900, 2015.

[15] G. Ivosev, L. Burton, and R. Bonner, “Dimensionality reduction and visualization in principal component analysis,” *Anal. Chem.*, vol. 80, no. 13, pp. 4933–4944, 2008.

[16] X. Liu, “Discriminative Principal Component Analysis for High Dimensional Classification with Applications in NIR Spectroscopy,” Dissertation, University College London, 2021. [Online]. Available: https://discovery.ucl.ac.uk/id/eprint/10107869/1/Xiaoke\_Thesis\_0813.pdf

[17] A. Tharwat, T. Gaber, A. Ibrahim, and A. E. Hassanien, “Linear discriminant analysis: A detailed tutorial,” *AI Commun.*, vol. 30, no. 2, pp. 169–190, 2017.

[18] “ISLR Chapter 4: Classification (Part 2: Linear Discriminant Analysis),” Amit Rajan. Accessed: May 13, 2024. [Online]. Available: https://amitrajan012.github.io/post/classification\_part2/

[19] “10.3 - Linear Discriminant Analysis | STAT 505.” Accessed: May 13, 2024. [Online]. Available: https://online.stat.psu.edu/stat505/lesson/10/10.3

[20] D.-N. Jiang, L. Lu, J. Tao, and L.-H. Cai, “Music type classification by spectral contrast feature,” Feb. 2002, pp. 113–116 vol.1. doi: 10.1109/ICME.2002.1035731.