Laboratory assignment

Component 4 - Related Work

Authors: Leordean Ada Alexandra, Selegean Victor
Group: 242

January 14, 2026

1 Supervised Learning

1.1 Related Work

Machine learning techniques have been widely applied in the field of music information retrieval to analyze, classify, and predict musical properties using audio features. Publicly available datasets containing Spotify audio descriptors have enabled the development of supervised learning models for tasks such as popularity prediction, genre classification, and mood detection.

1.2 Music Classification Using Audio Features

Several studies address supervised classification problems using low-level and high-level audio features similar to those provided by Spotify, including danceability, energy, tempo, and valence. In [SGU16], the authors present an overview of machine learning approaches for music analysis, highlighting that distance-based classifiers such as k-Nearest Neighbors (k-NN) are frequently used as baseline methods due to their simplicity and effectiveness in feature-based classification tasks.

Bogdanov et al. [BHOB19] explore the use of Spotify audio features for music recommendation and categorization. Their work shows that simple classifiers, including k-NN, can achieve competitive accuracy when audio features are properly normalized. They emphasize that k-NN is particularly suitable when feature similarity reflects perceptual similarity between songs.

1.3 k-Nearest Neighbors in Music Classification

The k-Nearest Neighbors algorithm has been successfully applied in music classification tasks such as genre recognition and popularity categorization. In [TC02], k-NN is used to classify music tracks based on extracted audio features, demonstrating that distance-based methods can effectively separate classes in the feature space.

More recent studies confirm that k-NN performs well on tabular music datasets when combined with feature scaling and appropriate distance metrics. Typical performance evaluation is conducted using classification accuracy, precision, recall, and F1-score. Reported classification accuracies for binary music classification tasks often range between 65% and 80%, depending on the number of features and the value of k.

1.4 Comparison with the Present Work

In this project, a supervised classification approach based on the k-Nearest Neighbors algorithm is employed to classify Spotify tracks according to their popularity category. The input consists of normalized Spotify audio features, and class labels are derived from the popularity score.

The use of k-NN aligns with established practices in the literature, where distance-based classifiers are commonly adopted for music classification problems. Performance is evaluated using standard classification metrics, enabling direct comparison with previously reported results. Although k-NN is a relatively simple method, it provides a strong baseline and offers intuitive interpretability by relying on similarity between songs in the feature space.

2 Unsupervised Learning

The problem of clustering songs based on genre is one that emerged in the face of Music Information Retrieval systems. As streaming platforms like Spotify and YouTube emerged and became the mainstream way of consuming music, so has the need for classifying a large quantity of unlabeled tracks. Casting aside the possibility of human labeling leaves the option of unsupervised clusering.

2.1 Defining a genre

The main issue faced by others attempting a similar project is the fact that "genre" is a cultural label. As such, it may not always be in-sync with the numerical features of the track it is associated with. Even more, two tracks with very similar numerical features may belong to different genres, depending on context. The differences may be more subtle and refer to lyrical content, or the year when a track was released (a pop track from the '80s will sound very different to a pop track from the 2020s).

Genres are also subjective. The literature suggests that humans agreement on genre classification caps at around 80%, fixing a theoretical ceiling to any algorithm that may try to complete the same task[FG16].

One advantage of automated classifiers is that they treat the input space of songs as a continuous topology, instead of a set of "buckets" with an associated taxonomy. This, in theory, should allow the algorithms to accommodate tracks that span the border regions of a genre cluster.

2.2 Features

Within our dataset, we are working with high level audio features like "danceability" and "energy". This is a significant departure from previous studies ([KYK15], [Bah18]) where the algorithms worked with low level audio features such as Mel-Frequency Cepstral Coefficients, Chroma Features, and Linear Prediction Coefficients. Higher level features are derived through different techniques which aim to estimate a given track's subjective "feel", rather than its raw signal properties.

2.3 Previous Results

In this project, K-Means Clustering will be applied to a 13-dimensional feature space, with K being set to either 35 (total genre count), or 21 (total genre count with more that 1% share of the track count).

In 2015, Kim, Yun, and Kim [KYK15] compared the clustering performance on low level feature sets using a K value of 3 and 5. For 3 genres, the model achieved a purity of 84.4%, while for 5 genres, the purity dropped to only 62%.

In 2023, in a Kaggle post [Clu23], Laila Masoud showed through the elbow method that there seems to be performance to be gained by going over 10 for K. Even still, most of the models' performance lies between K=2 and K=3.

2.4 Challenges

In order to tackle the issue of clustering songs using a dataset with high level features, the following challenges must be addressed:

- The disconnect between the low level signal features that can be extracted by a machine and the high level features identified by humans;
- The sparseness of the feature space, owing to the large number of features;
- Boundary tracks and hybrid genres lacking a definitive "correct" label;

References

[Bah18] Hareesh Bahuleyan. Music genre classification using machine learning techniques, 2018.

[BHOB19] Dmitry Bogdanov, Perfecto Herrera, and Fernando Orduna-Bustamante. Music recommendation and discovery with spotify audio features. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 2019.

[Clu23] Spotify Songs Clustering. Laila masoud, 2023.

[FG16] Arthur Flexer and Thomas Grill. The problem of limited inter-rater agreement in modelling music similarity. *J New Music Res*, 45(3):239–251, July 2016.

[KYK15] Kyuwon Kim, Wonjin Yun, and Rick Kim. Clustering music by genres using supervised and unsupervised algorithms. In *Clustering Music by Genres Using Supervised and Unsupervised Algorithms*, 2015.

[SGU16] Markus Schedl, Emilia Gomez, and Julian Urbano. Music information retrieval: Recent developments and applications. *Foundations and Trends in Information Retrieval*, 8(2):127–261, 2016.

[TC02] George Tzanetakis and Perry Cook. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, 2002.