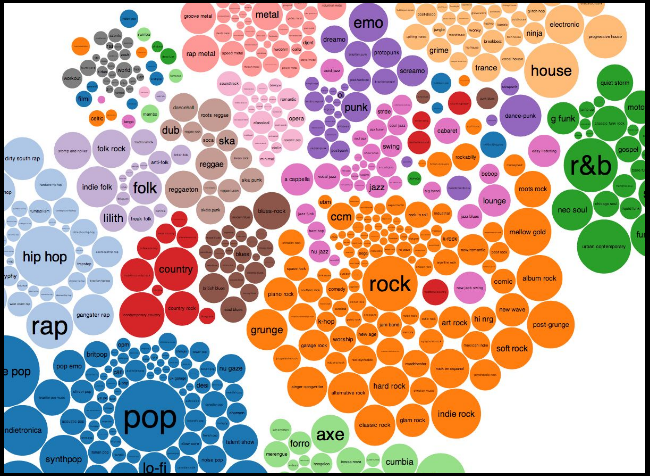
Exploratory Data Analysis of Various Musical Genres

George Hagopian & Tyler McKean
EECE 5644 Intro to Machine Learning

Table of Content

EDA Objective & GTZAN Dataset

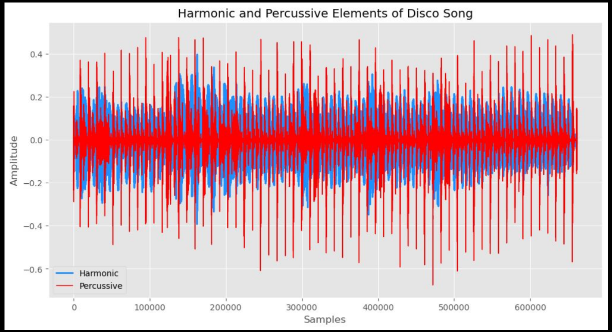
- GTZAN Dataset:
  - 1000 audio files from 10 different genres, each 30s long.
  - CSV file containing the different features of each audio file.
- EDA Goals:
  - Determine which classification model yields the best performance.
  - Determine which features are most useful for classification.
  - Relate these features to the different properties of musical genres.

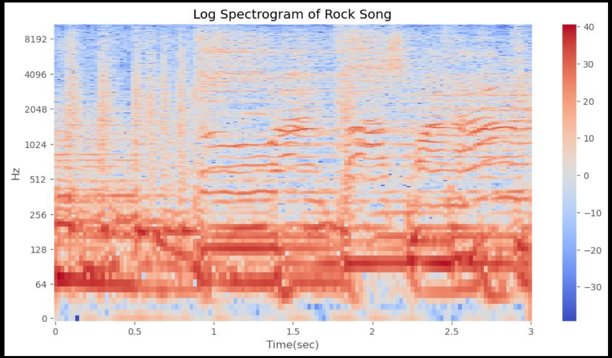


Clusters of Music Genres [9]

Features of Audio

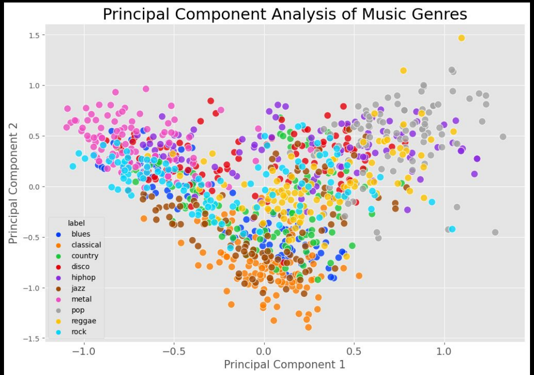
- 57 Features in Dataset
  - Mono .wav audio files
  - Sample Rate: 22,050 Hz
- Extracted using Librosa Module
  - Time-Domain
    - Tempo
    - Harmonic/Percussive Elements
  - Frequency-Domain
    - Log Spectrum
    - Bandwidth
    - Spectral Centroid
    - Chroma (Notes A - G)
- Implement methods of Dimensionality Reduction and Feature Selection to gain knowledge and aid classifiers

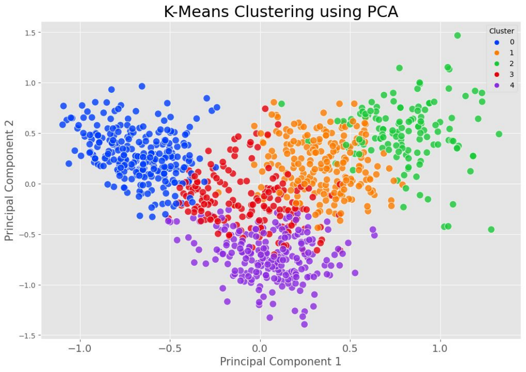




Dimensionality Reduction

- Principal Component Analysis
  - Reduced 57 Features down to 2 Principal Components
  - Tradeoff in dimension reduction to provide intuitive visualization for separation of genres
- K-Means Clustering
  - Genre labeling is abstract
  - Unsupervised approach to gain understanding of how similar some genres are
  - Reduced 10 genres into 5 clusters after applying PCA





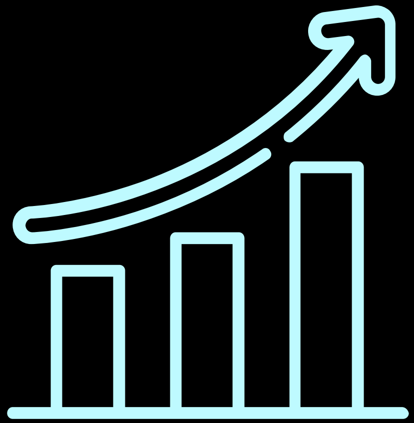
Initial Feature Selection

- Basic Feature selection:
  - ran Naive Bayes using a single feature to classify between all 10 genres
  - able to obtain an initial ranking by iterating through features
  - Average accuracy: 0.201
- Issues:
  - Only evaluates 1 feature at a time
  - Assumes independence, (generally not true)

Top 5 features (single-feature classification)	
Feature:	Accuracy:
spectral_bandwidth_mean	0.283
chroma_stft_mean	0.280
rolloff_mean	0.269
perceptr_var	0.267
spectral_centroid_mean	0.245

Sequential Feature Selection

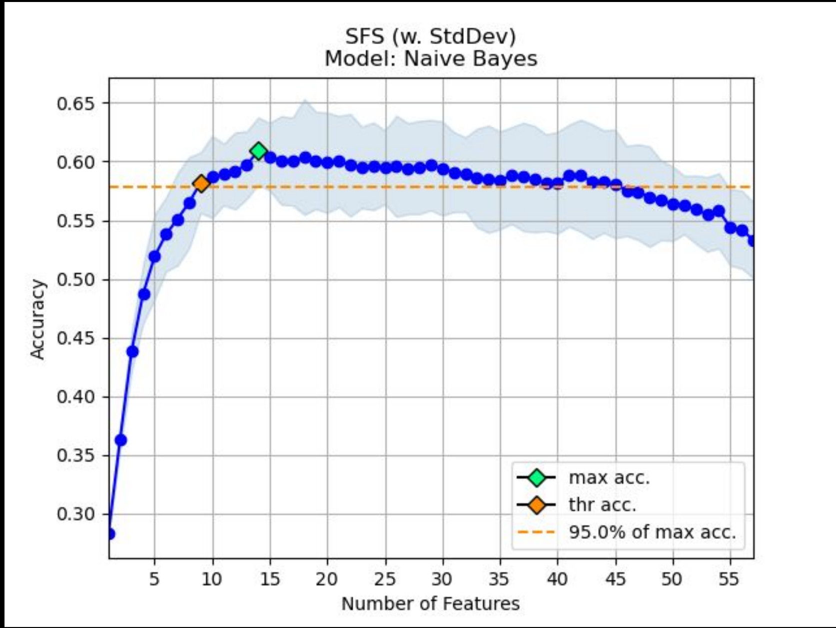
- Expands on our initial, basic feature selection:
  - Start with 0 features
  - adds the feature that gives the best performance
  - Stop when we hit a specified # of features.
- Allows us to get a sense for the most important features, as well as performance vs # of features
- Drawbacks:
  - Not as efficient as PCA
  - Computationally intensive



Sequential Feature Selection

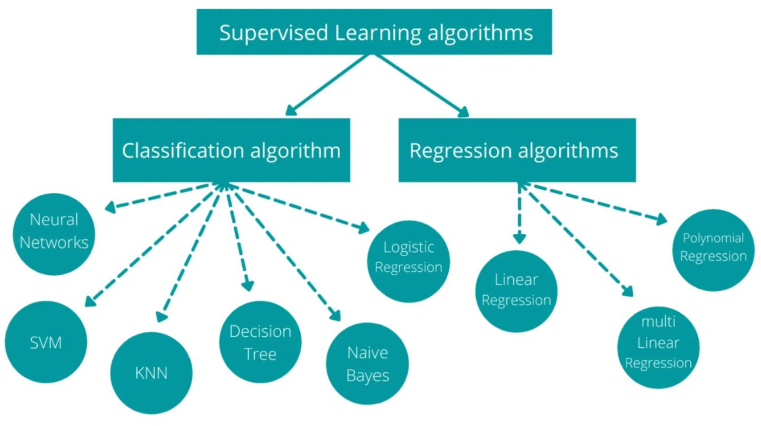
Example using Naive Bayes:

- More features - not necessarily better!
    - Accuracy using all features: 0.533
    - Max accuracy: 0.609, at 14 features
    - ~95% of max. accuracy: 0.582, at 9 features



Model Selection

- Known Labels = Supervised Learning!
- Key Aspects for Model Selection:
  - Performance
  - Complexity
  - Data Size/Dimension
  - Implementation
- From PCA and Feature Selection:
  - Classification is Linear Problem
  - Models with High Bias/Low Variance
  - Easy to implement, but also could require fine tuning
- Classification Models Chosen:
  - Logistic Regression
  - Linear Support Vector Classification
  - K-Nearest Neighbors



Overview of Supervised Machine Learning Algorithms [8]

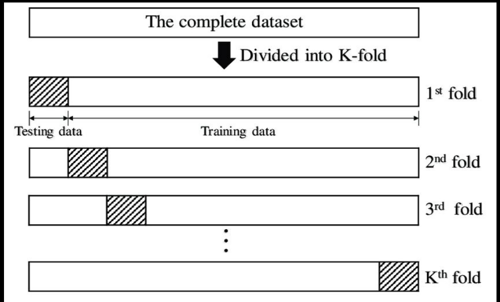
Optimization Methods

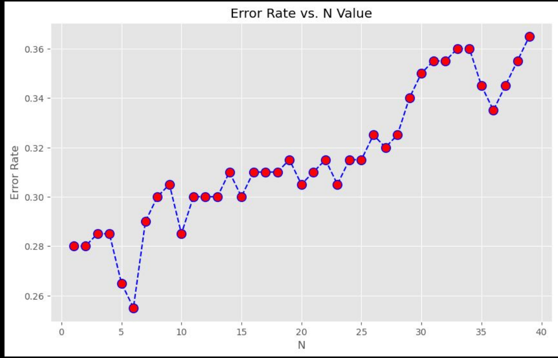
Feature Selection/Reduction:

- Previously mentioned

Optimal Methods for Training/Testing

- Stratifying training/test samples
  - 10% of test/training for each genre
- K-Fold Cross Validation
  - Train models with K folds
  - Compute mean accuracy performance
- Optimal Value *N* for KNN Model
  - Compute Error Rate for varying values of *N* - Neighbors
  - Use value *N* that minimizes KNN model error -> N = 6

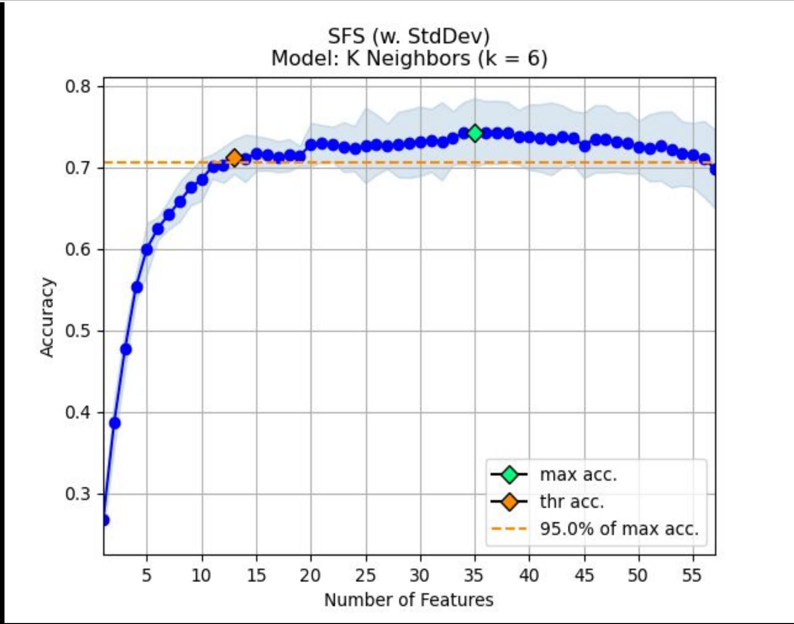




Classification Results

Model	Baseline accuracy (all 57 features)	Max accuracy	Threshold accuracy
Naïve Bayes	0.533	0.609 (14 features)	0.582 (9 features)
SVM (Linear)	0.687	0.716 (42 features)	0.681 (26 features)
Logistic Regression	0.661	0.685 (45 features)	0.651 (21 features)
K Nearest Neighbors (N = 6)	0.698	0.743 (35 features)	0.713, (13 features)

Classification Results



Sequential Feature Selection Results for K Nearest Neighbors

Feature Results

- For each classifier, we checked the features that SFS selected to give max performance.
- Several features were selected for all 4 classifiers:

Most consistent features:
Chroma_stft_mean
Chroma_stft_var
Mfcc10_var
Mfcc4_mean
Mfcc5_var
Mfcc9_mean
rms_mean
rms_var

EDA Considerations

Satisfactory Models
- Decision Trees
- Random Forest

Feature Selection/Reduction
- Lasso Regression: accuracy range: ~50%-60%

Further Improvements
- Neural Networks
  - Traditionally trained with 2D data
  - Methods of using Spectrograms of Audio previously done
- Increasing Dataset
  - Manually add more .wav files to dataset
  - Splicing audio into smaller clips

References

1. G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," in *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293-302, July 2002, doi: 10.1109/TSA.2002.800560.
2. L. K. Puppala, S. S. R. Muvva, S. R. Chinige and P. S. Rajendran, "A Novel Music Genre Classification Using Convolutional Neural Network," *2021 6th International Conference on Communication and Electronics Systems (ICCES)*, 2021, pp. 1246-1249, doi: 10.1109/ICCES51350.2021.9489022.
3. Changsheng Xu, N. C. Maddage, Xi Shao, Fang Cao and Qi Tian, "Musical genre classification using support vector machines," 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03)., 2003, pp. V-429, doi: 10.1109/ICASSP.2003.1199998.
4. Venturott, P. H. G. (2021, March 3). *Predicting Music Genres Using Waveform Features*. Medium. Retrieved December 7, 2022, from https://towardsdatascience.com/predicting-music-genres-using-waveform-features-5080e788eb64
5. Doshi, K. (2021, May 21). *Audio Deep Learning Made Simple: Sound Classification, Step-By-Step*. Medium. Retrieved December 7, 2022, from https://towardsdatascience.com/audio-deep-learning-made-simple-sound-classification-step-by-step-cebc936bbe5
6. Gandhi, S. K. (2021, July 9). *Finding Out Optimum Neighbours (N) Number in the KNN Classification using Python*. Medium. Retrieved December 7, 2022, from https://medium.com/analytics-vidhya/finding-out-optimum-neighbours-n-number-in-the-knn-classification-using-python-9bdcfefff58c
7. Pramoditha, R. (2020, December 20). *K-Fold Cross-Validation Explained in Plain English*. Medium. Retrieved December 7, 2022, from https://towardsdatascience.com/k-fold-cross-validation-explained-in-plain-english-659e33c0bc0
8. Doshi, K. (2021, May 21). *Audio Deep Learning Made Simple: Sound Classification, Step-By-Step*. Medium. Retrieved December 7, 2022, from https://towardsdatascience.com/audio-deep-learning-made-simple-sound-classification-step-by-step-cebc936bbe5
9. *Music Popcorn – A Visualization of the Music Genre Space*. Music Machinery. (2013, September 23). Retrieved December 12, 2022, from https://musicmachinery.com/2013/09/22/5025/
10. S. Raschka, “Sequentialfeatureselector: The popular forward and backward feature selection approaches (including floating variants),”- mlxtend. http://rasbt.github.io/mlxtend/user_guide/feature_selection/SequentialFeatureSelector/#sequentialfeatureselector-the-popular-forward-and-backward-feature-selection-approaches-including-floating-variants.
11. J. Lyons, “Mel Frequency Cepstral Coefficient (MFCC) tutorial,’’ Practical Cryptography. [Online]. Available: http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/. [Accessed: 13-Dec-2022].

Thank you!