Music Genre Classification

By: Jacob and Josh

Our Goal

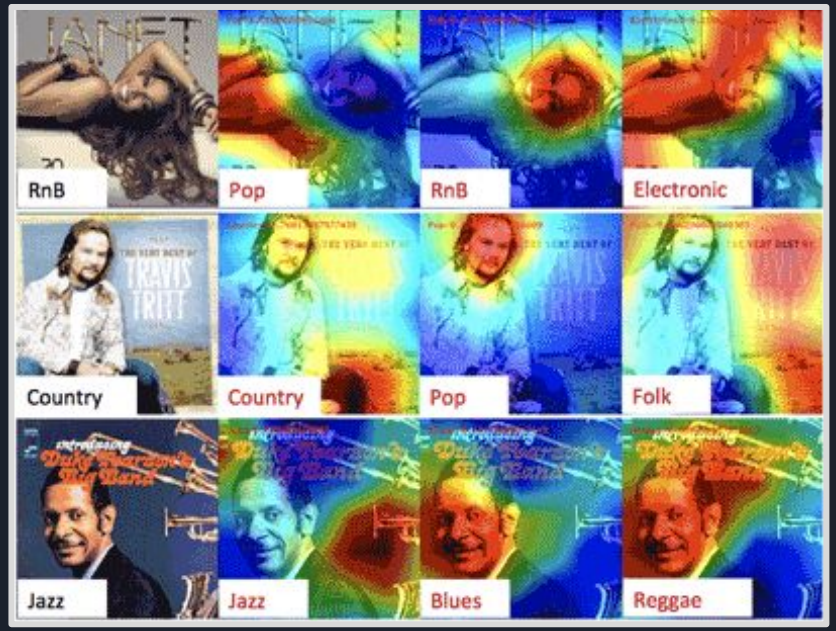
- To build a predictive model to classify the genre of a given song into 1 of 10 genres
  - Genre - the type or style of music
- Take a 30 second audio clip and classify its genre using our model

Importance of Music Genre Classification

- Helps music streaming platforms to…
  - Provide relevant recommendations
  - Curate playlists
  - Play relevant songs once our queue is over
  - More accurately label newer & or niche genres

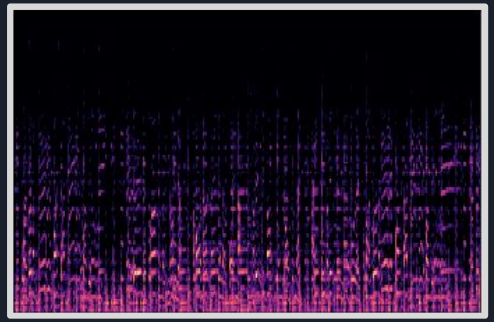
Similar Works

- Multimodal Deep Learning for Music Genre Classification
  - CNN
  - Multimodal approach
  - Single-label
  - Multi-label
  - Album cover



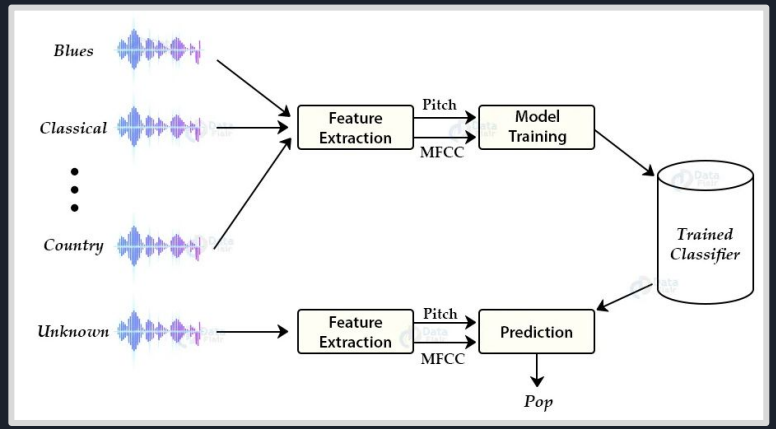
Data

- GTZAN Music Genre Dataset
- 1,000 song samples, each 30 seconds long, belonging to 10 different genres
- Contains
  - 3 second clip features
  - 30 second clip features
  - 30 second audio clips
  - Visual representations of the clips as Mel Spectrographs
- Challenges
  - Extracting proper/sufficient features
  - Restricted number of samples



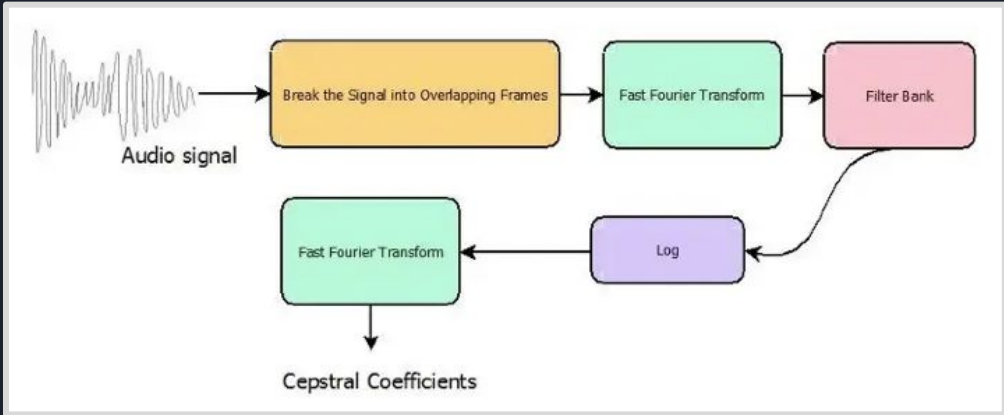
Approach

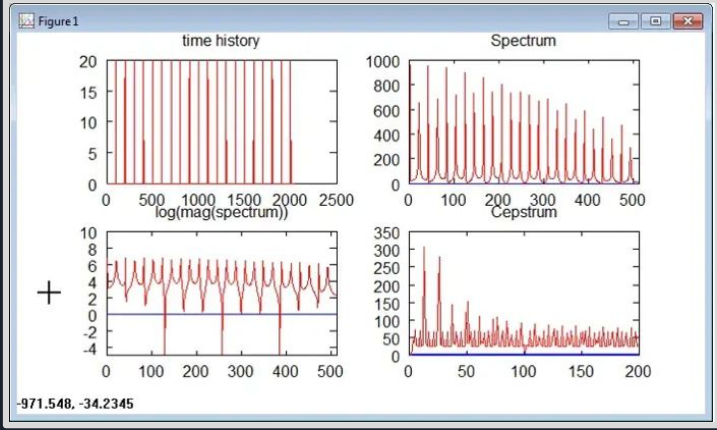
- Decided to go with K-nearest neighbors using MFCC's
- Feature extraction
- Mel Frequency Cepstral Coefficients
  - Divide signals into clips 20-40 ms long
  - Identify frequencies present in each clip
  - Separate linguistic frequencies from the noise
  - Use DCT to separate the noise
- Make predictions using KNN
- Try and match or beat the accuracy of past work

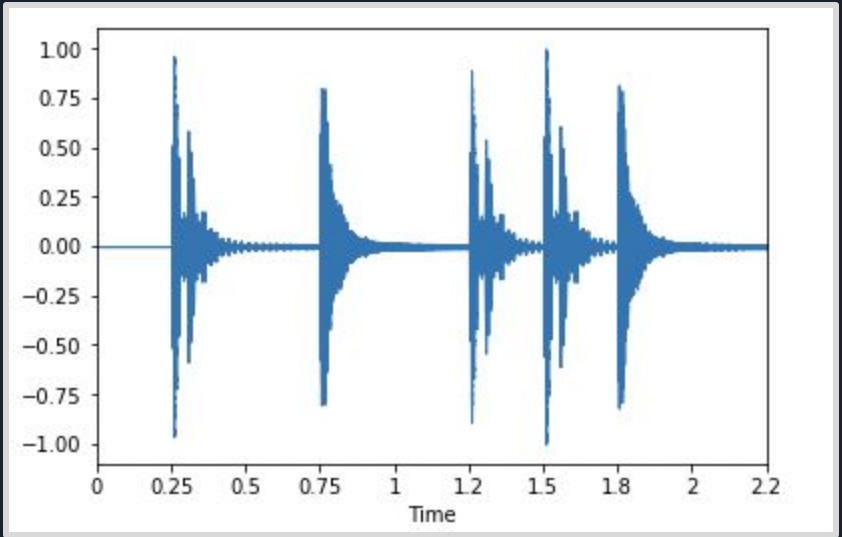


Mel Frequency Cepstral Coefficients (MFCC)

- Cepstrum - rate of change of spectral bands
- Fourier Spectrum
- Quefrequency domain -> cepstrum
- Mel scale - relates perceived frequency of tone to actual frequency
- Any sound generated by humans -> shape of vocal tract





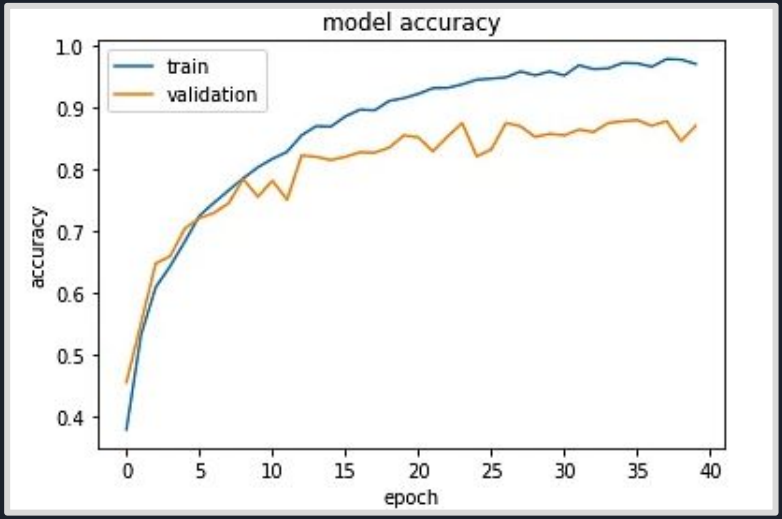

1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
0.25
0.5
0.75
1
1.2
1.5
1.8
2
2.2
Time

Process/Model

- Get distance between feature vectors and find neighbors
- Identify nearest neighbor
- Evaluate model
- Extract features from data (MFCC's)
- Train the model
- Make predictions using KNN
- Test with new untrained audio file

Results and Performance

- After many tweaks and iterations -> ~70% accuracy on test data
- Pretty good considering size of data and others' scores
- After researching - CNN seems to be more accurate using Mel Spectrographs
    - ~88% accuracy



CNN model accuracy

Conclusion

- In the future - more data could create a better model
- Try using CNN for better results
- Classify songs into more than just 10 genres - more useful/realistic