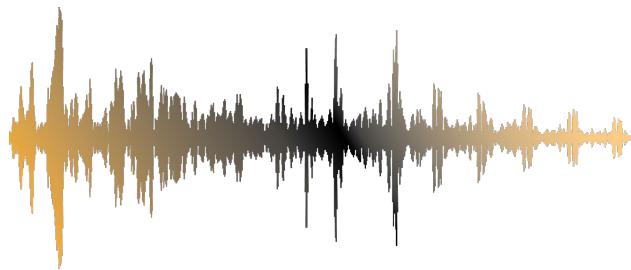

Predicting song popularity for different genres

— Kathy Li, Sa Qu, Rulan Xiao —

What affects song popularity for different genres?

- What factors make a song gain popularity?
 - Artist recognition
 - Marketing
 - Radio play
 - ...many more
- What about the actual **audio features** of the song?
- Do these factors differ **by genre**?
- If we can leverage audio features to predict song popularity by genre, we can predict award show outcomes and more



Overview of data

- We'll be using the the **Free Music Archive (FMA)**, which amasses:
 - Audio files of **106,574 tracks from 161 genres**
 - Track information, such as **artist, genre, and listen count**
 - Information about audio features, including **danceability, tempo, energy, loudness**
 - Available at <https://github.com/mdeff/fma>
- We may also extract further audio features from the audio files using **LibROSA**, a python package for music analysis

Data sample

artist

artist_mbtags: shape = (4,)

this artist received 4 tags on musicbrainz.org

artist_terms: shape = (12,)

this artist has 12 terms (tags) from The Echo Nest

danceability: 0.0

danceability measure of this song according to The Echo Nest (between 0 and 1, 0 => not analyzed)

duration: 211.69587

duration of the track in seconds

energy: 0.0

energy measure (not in the signal processing sense) according to The Echo Nest (between 0 and 1, 0 => not analyzed)

key: 1

estimation of the key the song is in by The Echo Nest

key_confidence: 0.324

confidence of the key estimation

loudness: -7.75

general loudness of the track

mode: 1

estimation of the mode the song is in by The Echo Nest

mode_confidence: 0.434

confidence of the mode estimation

What are we predicting?

- Quantify popularity: play count
- Different characteristics that make songs popular in different genre

E.g. What contributes most to the popularity of Jazz songs? Danceability? Energy? Or key? What contributes most to the popularity of Pop songs?

- Compare those things for different genre and see if there is any common factor.

Method

Clean data

Look for **redundancies and errors** in data; **normalize play count by genre**

Explore data

Perform **exploratory data analysis** and **select features**

Build model

Run **regression** on different genres to **predict popularity**

Enhance + adjust

Test **model performance** and add features / further analysis (i.e., lyrics)

How will we validate our model?

- We'll test against the **top Spotify songs dataset**,
<https://www.kaggle.com/nadintamer/top-tracks-of-2017/data>
- We'll run the model with Spotify data, and rank the songs based on predicted play counts
- Compare that to the real ranking and see if the model works

Questions / further ideas to explore

- Can we incorporate **sentiment analysis of lyrics**?
- Can we further analyze audio files to **extract more features** beyond what's already in the dataset?
- Can we **compare songs in our dataset to Grammy-winning songs** to discern which factors are predictive of award-winning potential?
- Can we write a program that **generates popular songs**?

