Topological Data Analysis on Music Data

Team Members: **Siddhant Ranade** – u1141131, Yash Gangrade – u1143811

April 29, 2019

Background and Motivation

- Music in general is rich in structures. Songs can be viewed as a sequence of notes.

- This sequence of notes has repeating patterns, for instance, verses will often have exactly the same tune. Observing carefully, it can be observed that certain phrases (or shorter sequences of notes) occur more frequently than others.

Overview

- At a high level:
  - Define a way to convert a song to a point cloud
  - Compute persistent homology
  - Use the persistence barcodes as features for machine learning tasks

Data

- MIDI Files -- record musical “events” (i.e. a musical note being played)
  - Which key?
  - Pressed or released?
  - How hard? (“velocity”)
  - And at what time?
- Complete description of the song
- Can be played back with a “sound font”
- A lot of music is available in this format

Point clouds (Embedding)

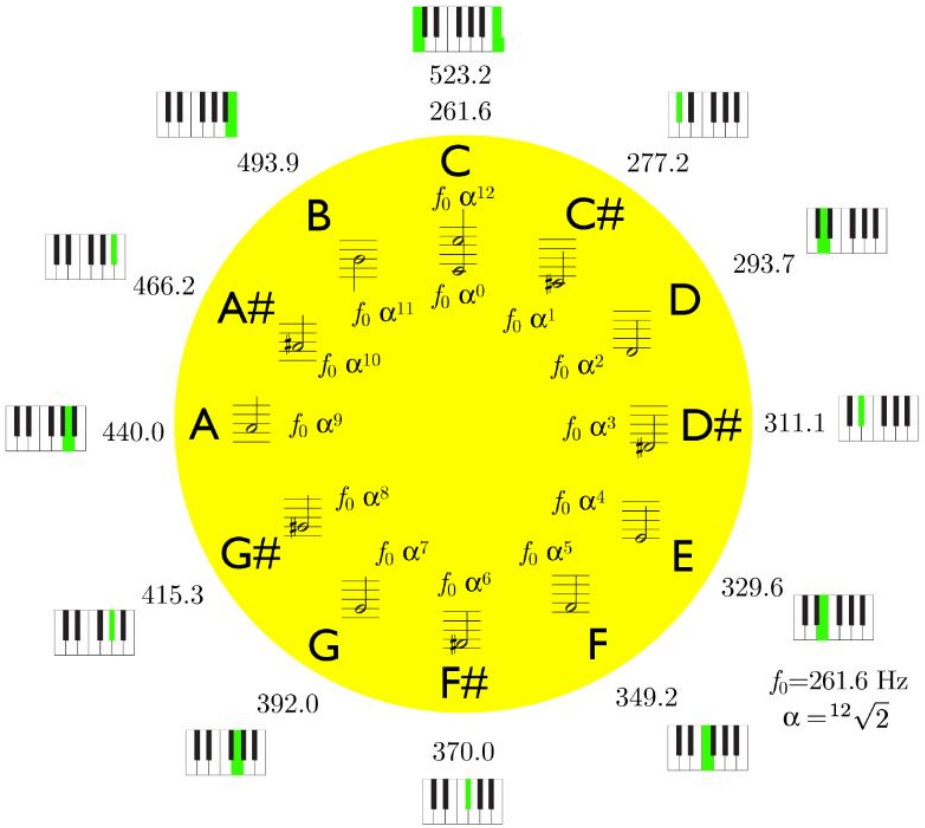


Fig. 1: Circle of Notes*

- Single note embedding -- 1D point cloud
- Notes can be perceived as lying on a circle. Distance between the notes is defined as the distance on this circle.
- Distance between C and F is 5, distance between C# and G is 6, and so on.
- Does not encode temporal information

*Source: Sethares et al., Topology of Musical Data

Point clouds (Embedding)

- Use a time series embedding:
- Consider (overlapping) sequences of consecutive notes with size d
- E.g. consider the song with notes E E F G G F E D C C D E E D D
- Its length-4 time-series embedding (a 4D pointcloud) would look like:
  - E E F G
  - E F G G
  - F G G F
  - …
  - E E D D
- Simple way to encode temporal information

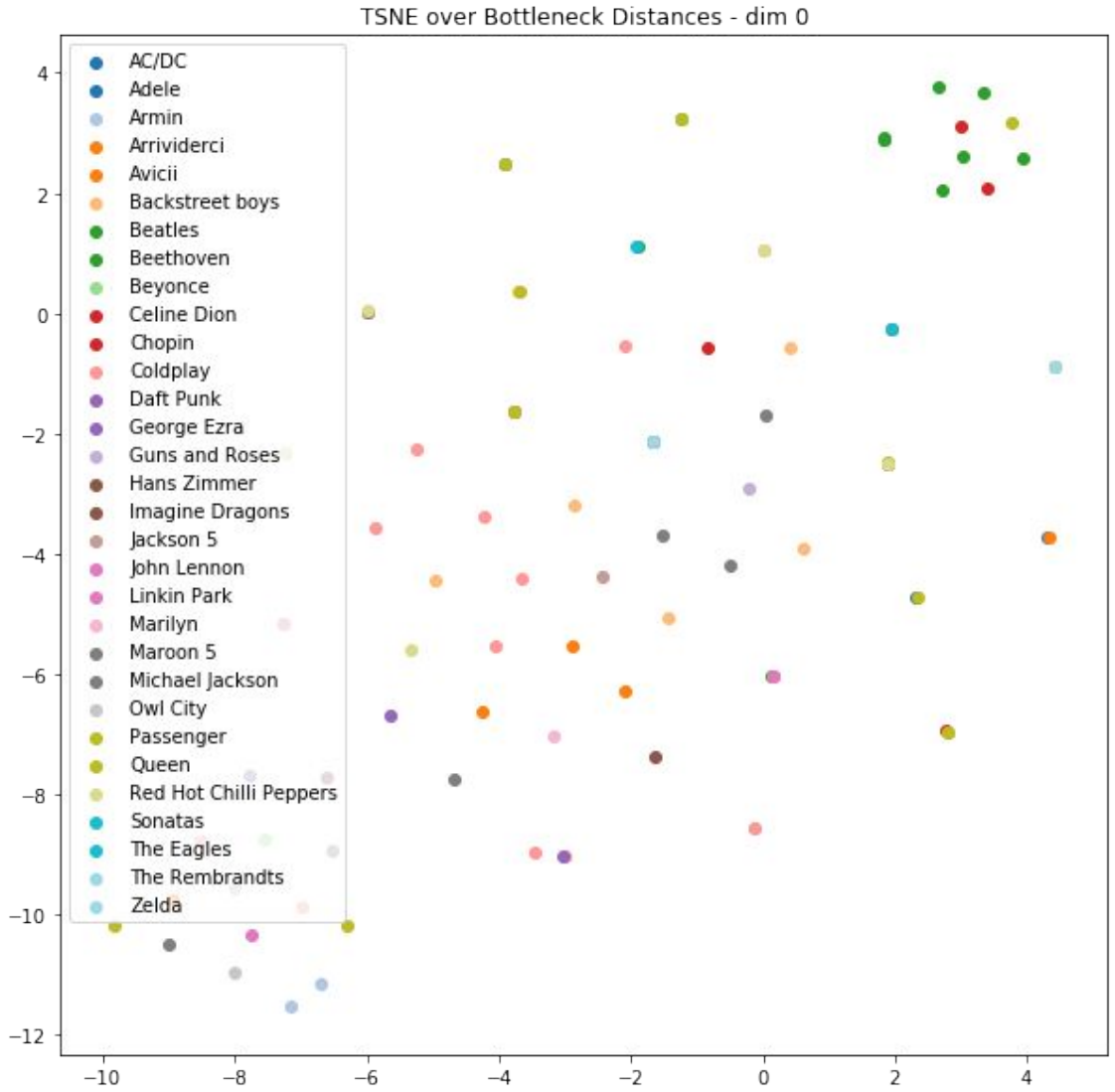
Time Series Embedding

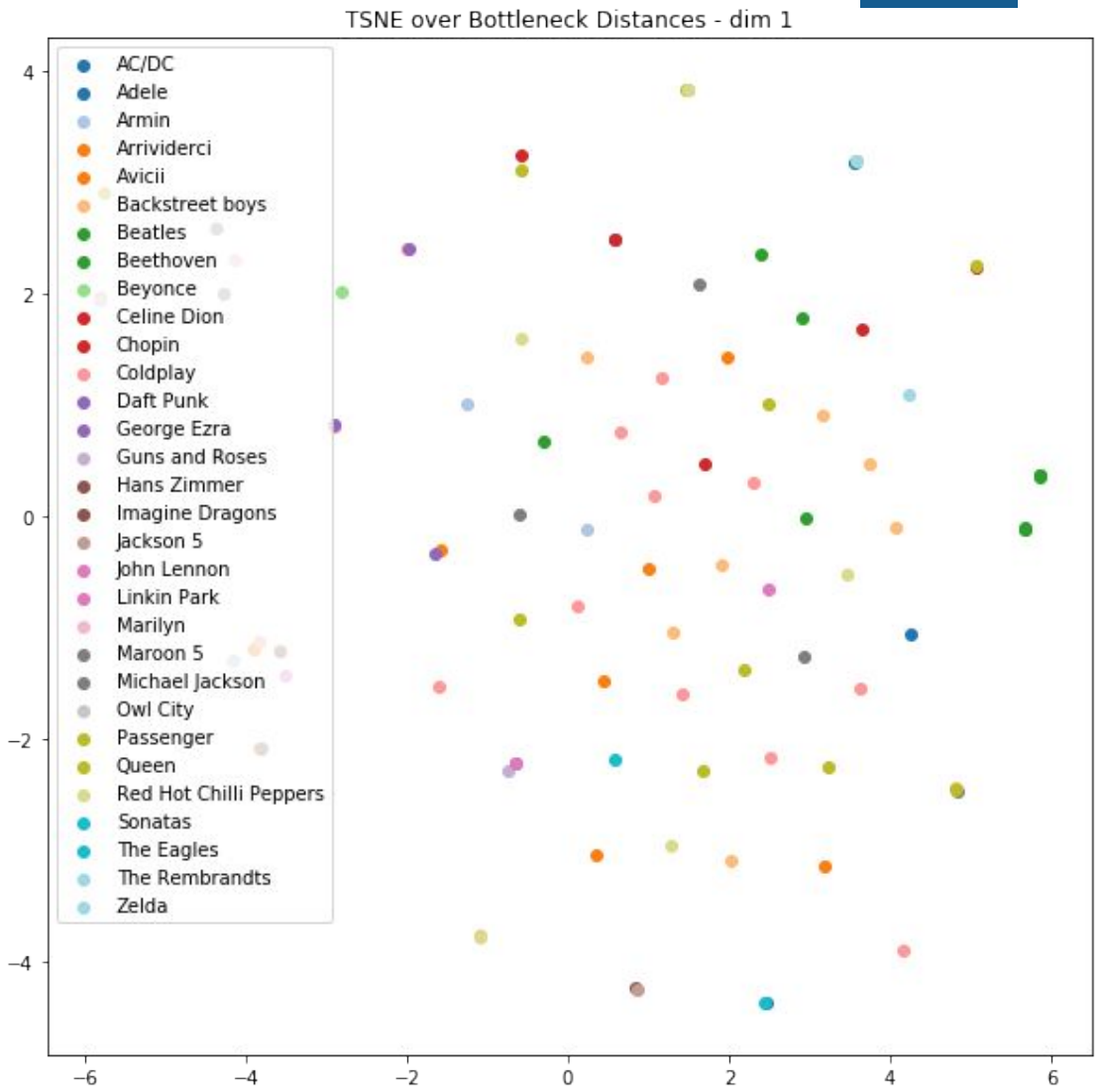
- The distance between two sequences (i.e. points of the pointcloud) is defined to be the sum of distances between corresponding notes
- E.g. the distance between (E, E, F, G) and (E, F, G, G) is 0 + 1 + 2 + 0

TDA+ML Tasks

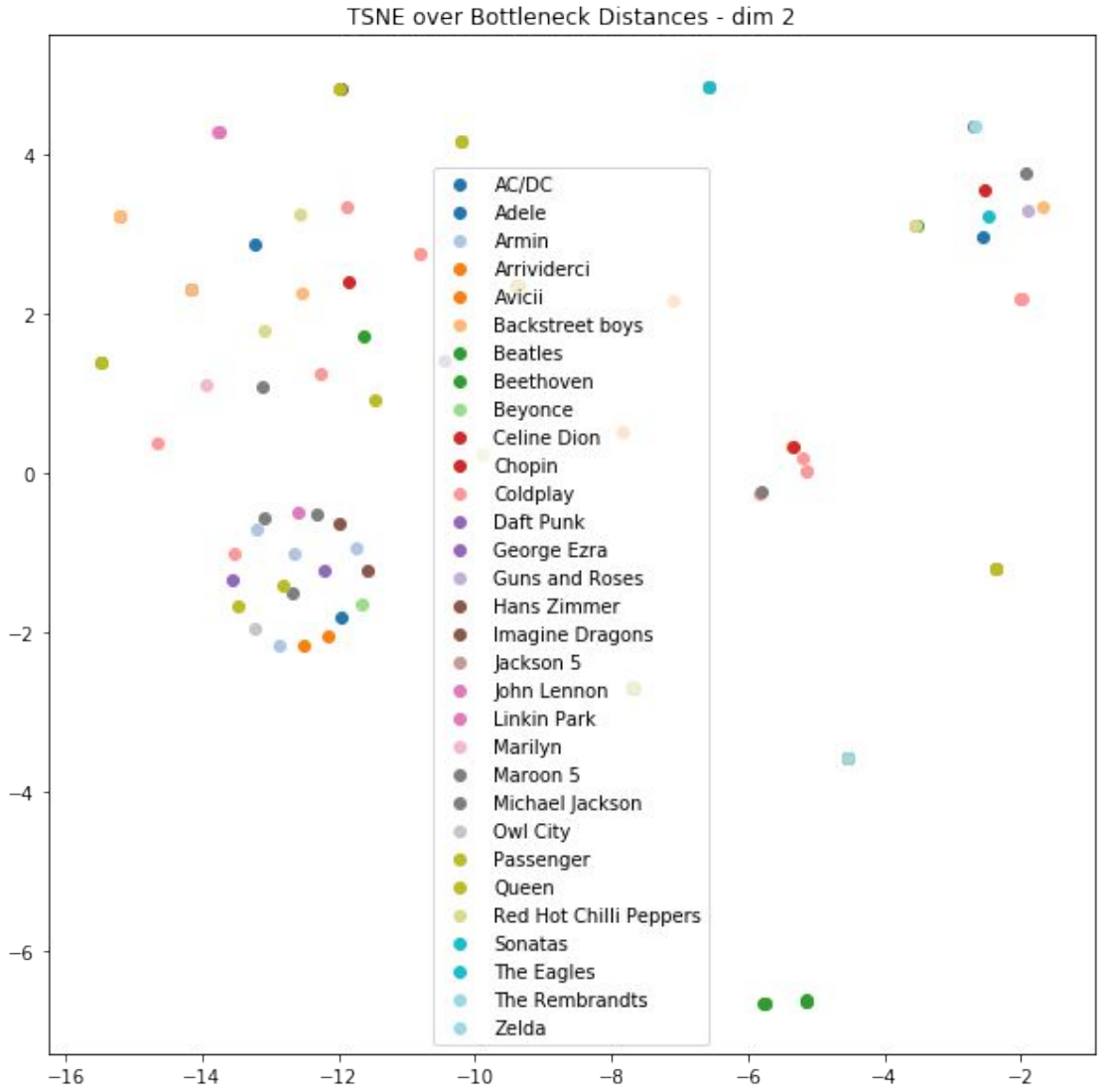
- Compute the persistence barcodes of the pointcloud.
- Compute pairwise Wasserstein and Bottleneck distances, embed the point-clouds into 2D using TSNE, and visualize that
- Train SVMs to perform tasks such as artist classification and genre classification.

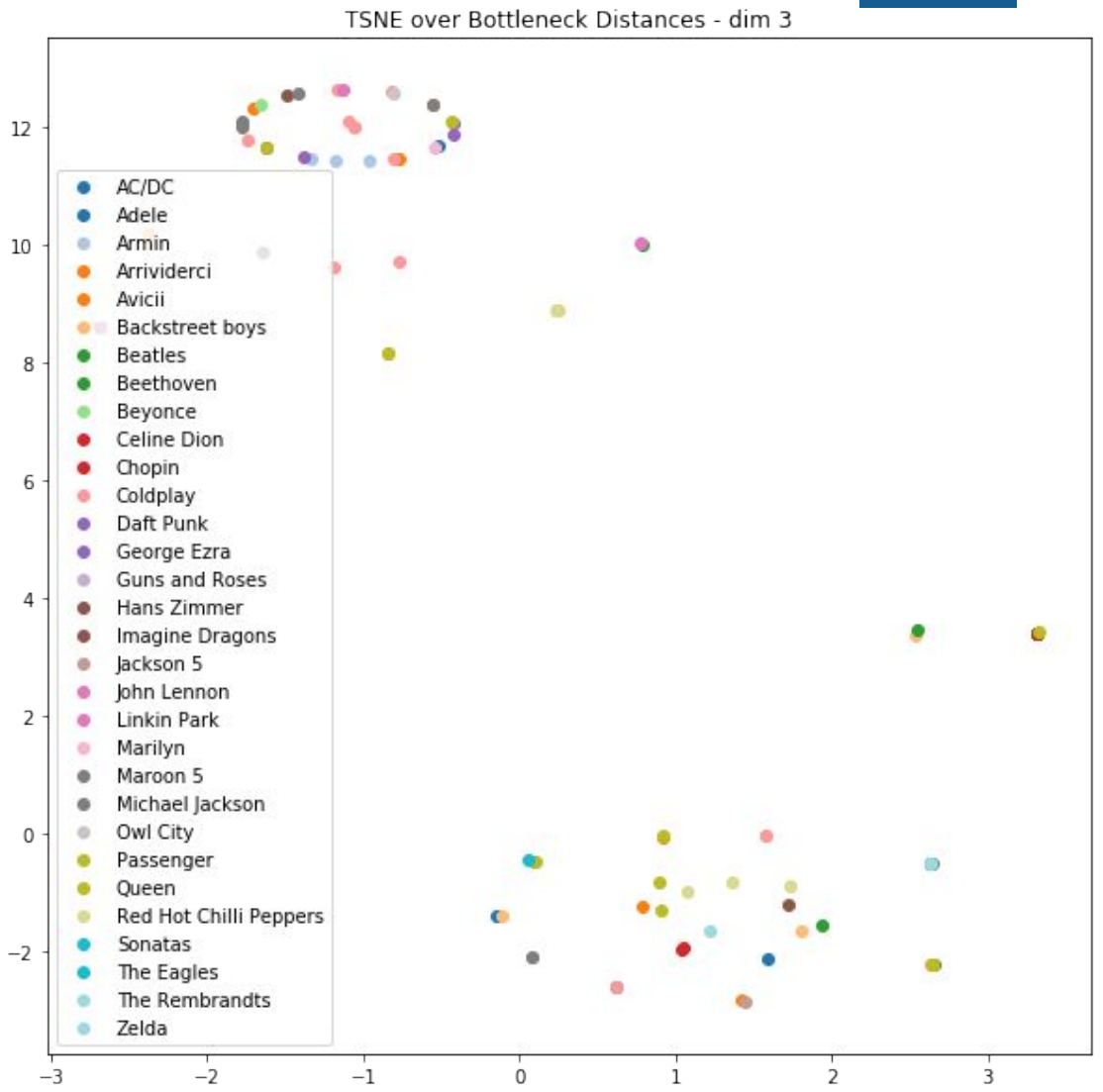
Results - Artist Identification



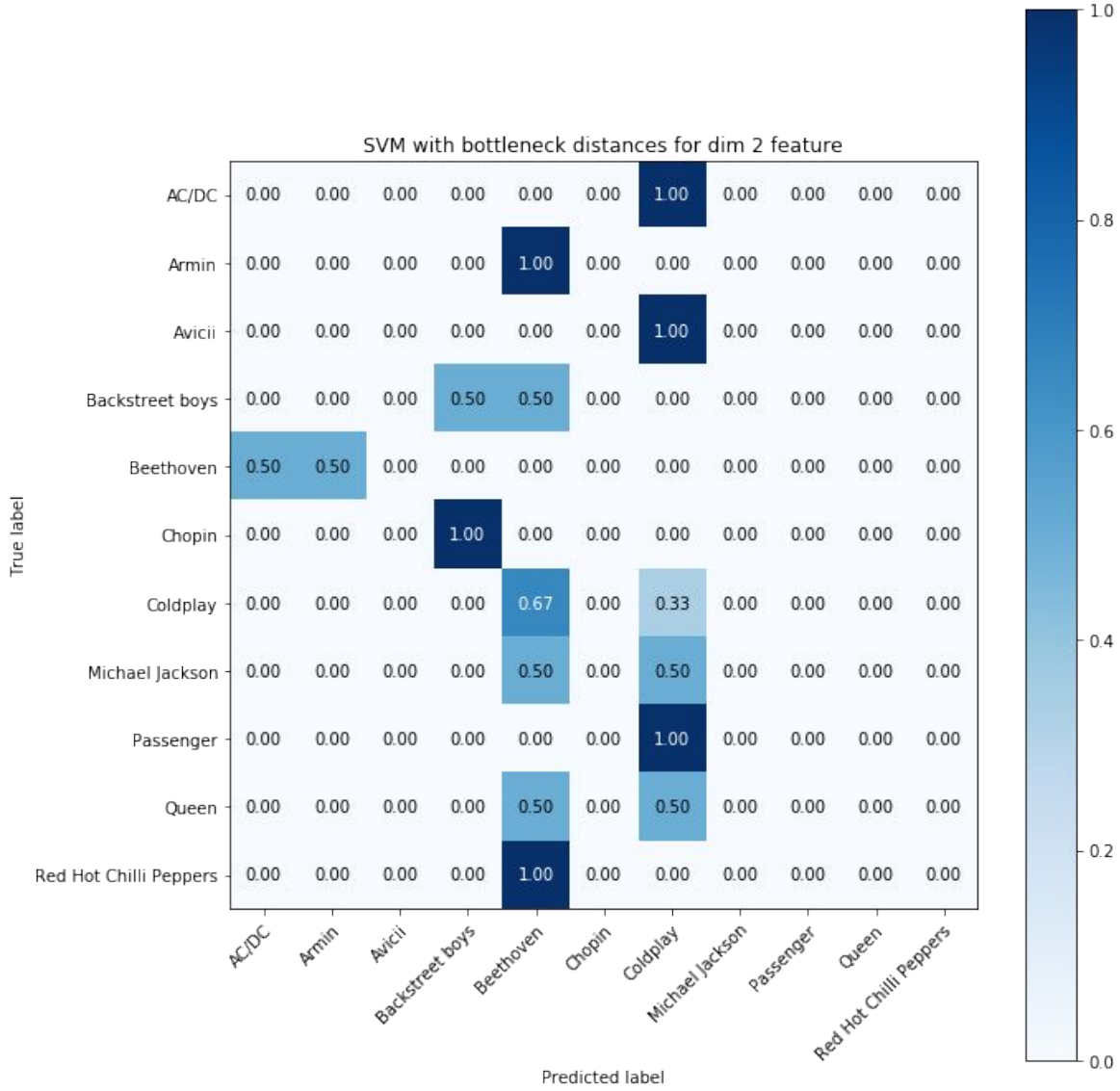


Results - Artist Identification



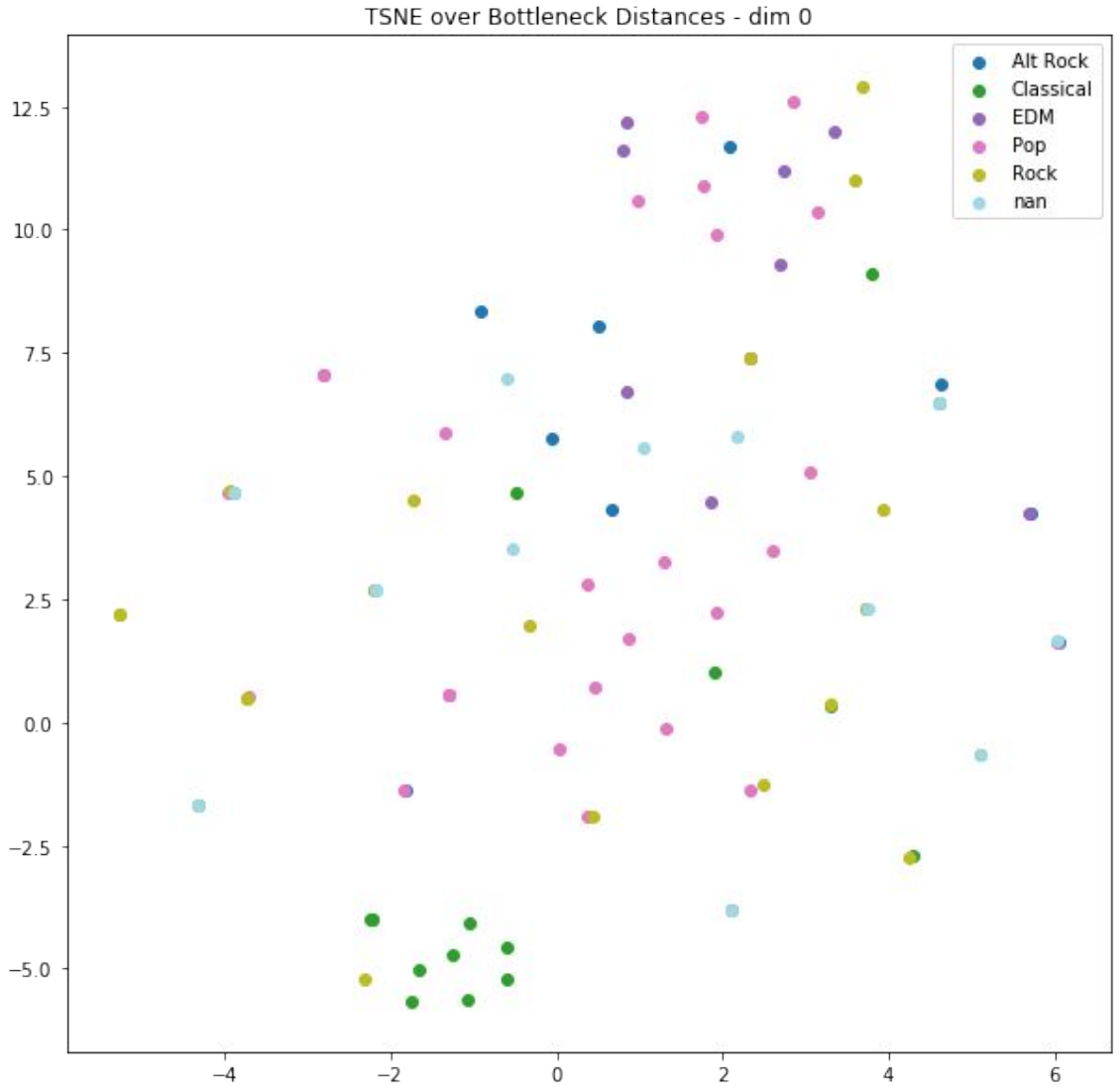


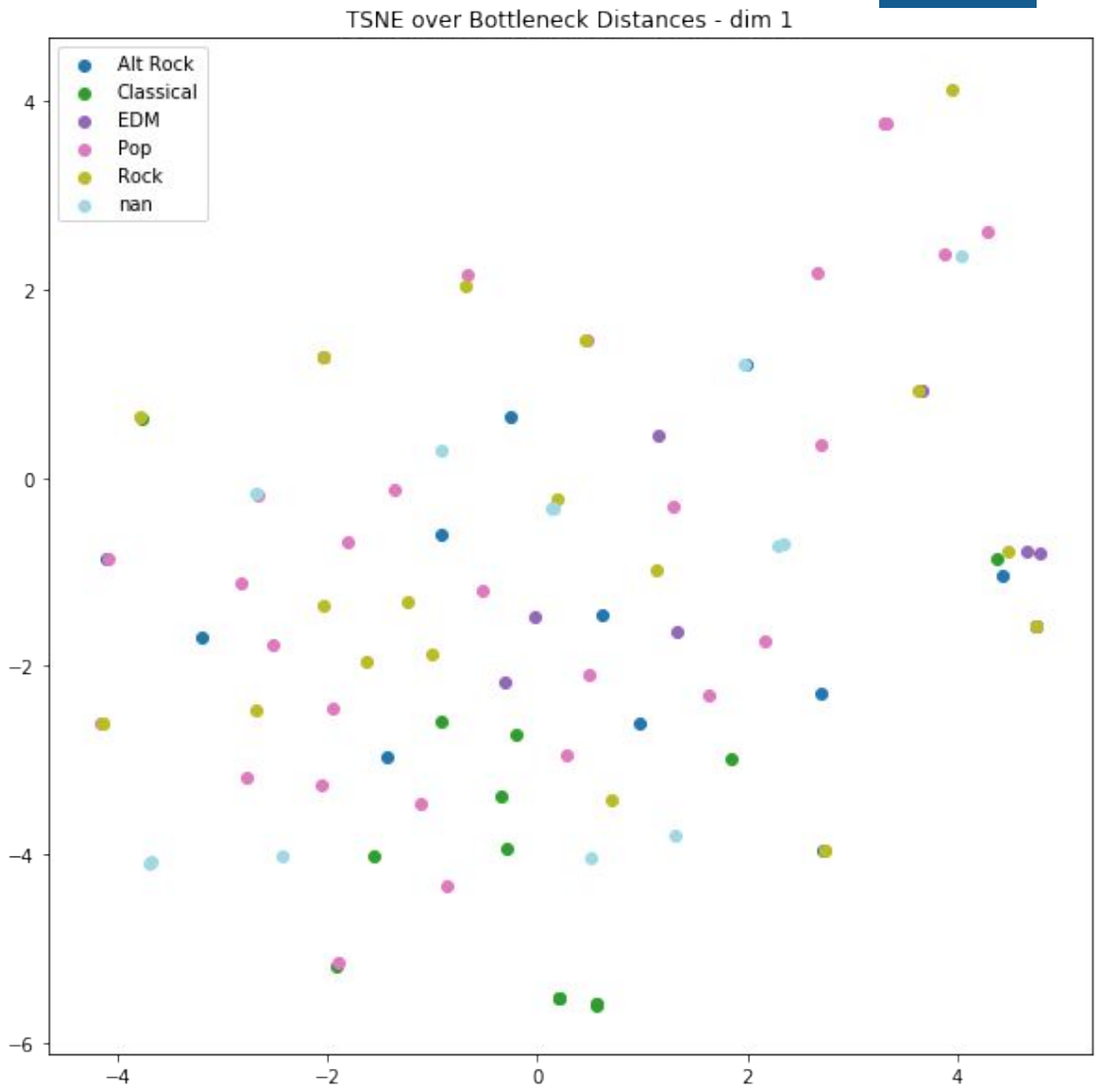
Results - Artist Identification



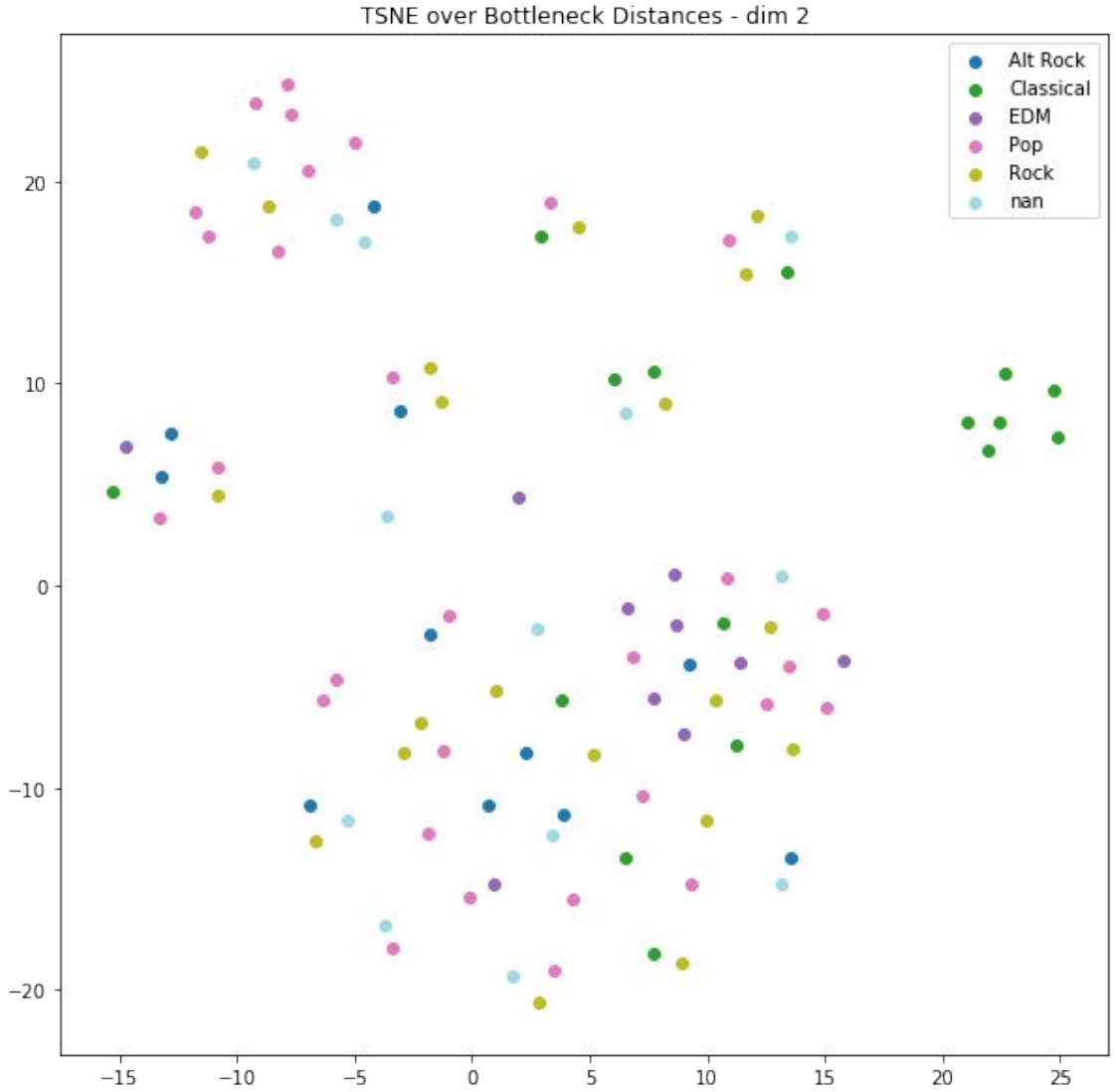
- 11 artists in test set
- Best accuracy = 17%
- Random chance would give us 9%

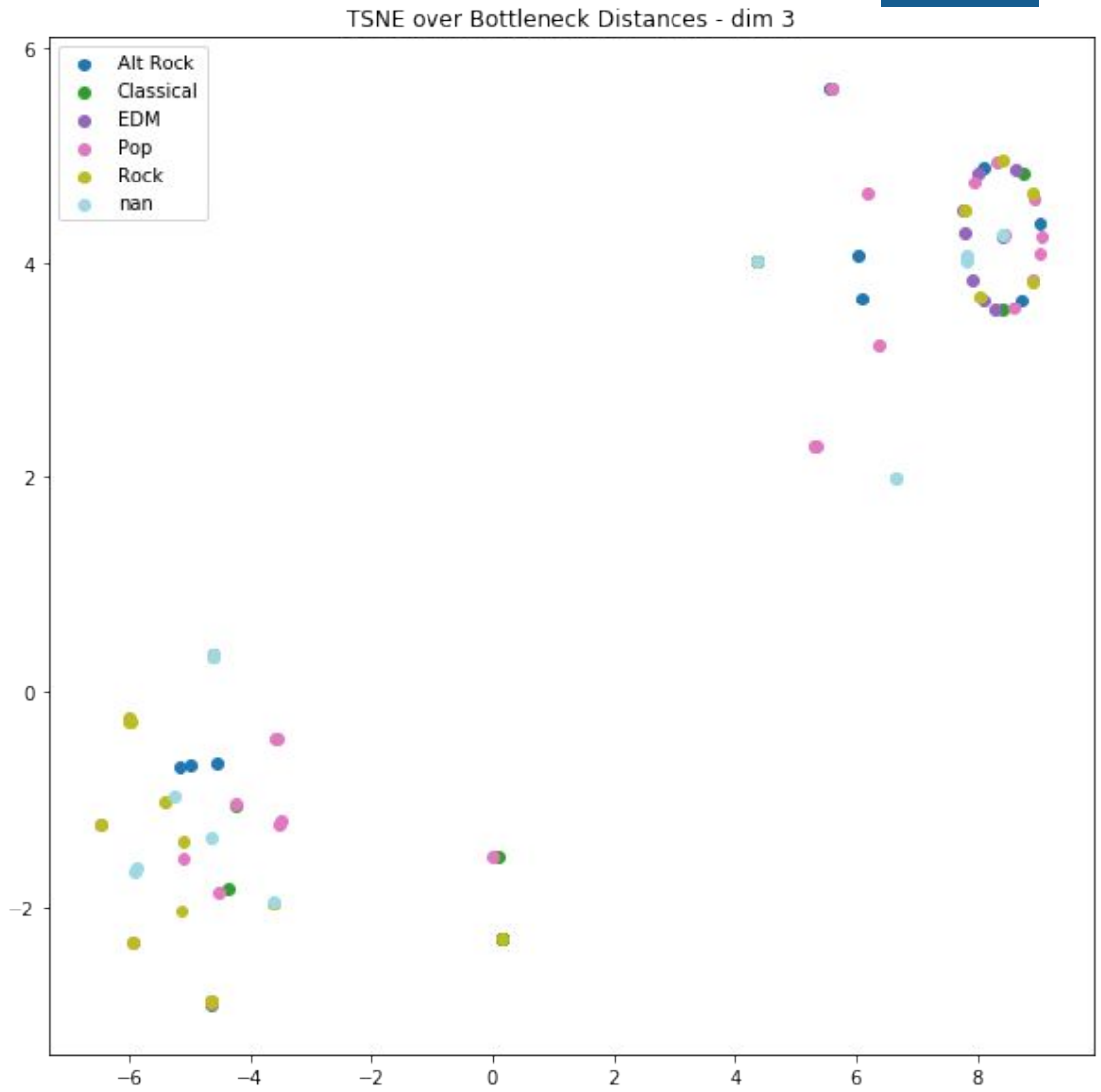
Results - Genre Classification



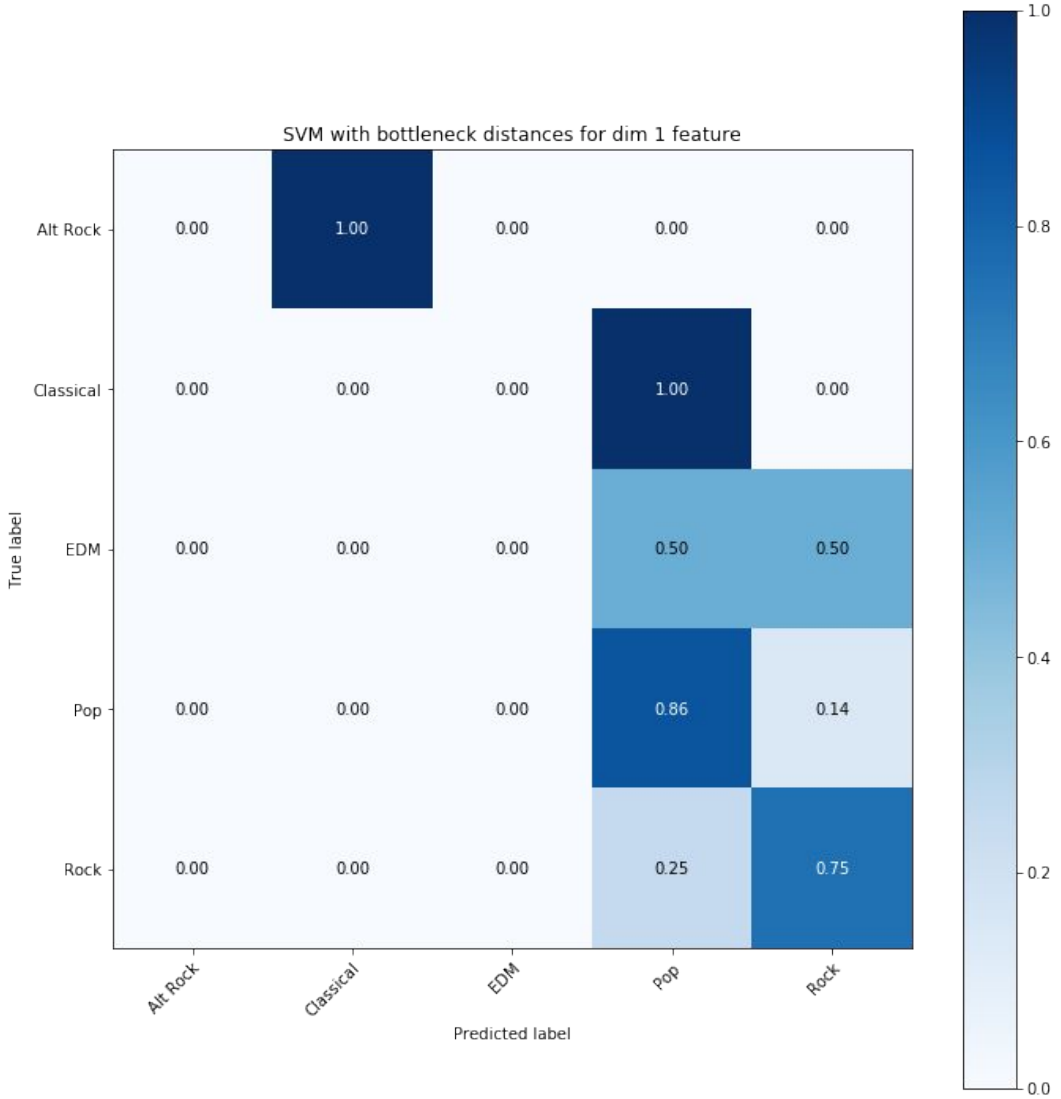


Results - Genre Classification





Results - Genre Classification



- 5 genres in test set
- Best accuracy = 52%
- Random chance would give us 20%

Conclusion, Future Work

- We use topological features to perform ML tasks on musical data.
- This works to a certain extent
- Collect more data (requires manual work at the moment) -- we have only 100 songs
- Use topological features in addition to the music data itself, perhaps with a better model.

Thank you!

Questions?