Overview

- Music composition, improvisation and accompaniment are creative acts that are challenging to capture algorithmically.
- Goal: Given the audio tracks of three instruments as input, generate an accompanying audio track for a fourth instrument.
- Approach: We formulated our problem as a classification problem, where each class is one of 128 possible note pitches.

Dataset

- We used the Lakh Dataset, which has over 170,000K MIDI files, and selected a subset of those files that have guitar, piano, strings, and bass tracks.

Methods

SVM

- As a simple baseline, we started with SVM. We compared performance on a linear and RBF kernel.
- Optimization Problem

\[
\min_{w, b, \zeta} \frac{1}{2} ||w||^2 + C \sum_{i=1}^{n} \zeta_i \quad \text{s.t.} \quad y^{(i)}(w^T x^{(i)} + b) \geq 1 - \zeta_i, \quad i = 1, \ldots, n, \quad \zeta_i \geq 0, \quad i = 1, \ldots, n
\]

- We address class imbalance problem by up-weighting positive examples.

CNN

- We address class imbalance problem by up-weighting positive examples.
- Our CNN implementation performs slightly better than SVM.

Results

- Both methods suffer from class imbalance problem, even with the use of up-weighting in the neural network loss function.
- Our CNN implementation performs slightly better than SVM.

Future Work

- Synthesize MIDI files to audio files and evaluate music with a Turing Test.
- Compare CNN performance to RNN and LSTM architectures and reframe the problem as seq-to-seq instead of classification.