

Enhancing Classical Music Generation with RNN

Alex Sokolov

Vishaal Prasad

Patrick Liu

Thomas Andy Keller

About this project

Motivations:

- Explore ways to improve RNN classical music generation
- Determine if objective measure of generated output exists

Data:

- ~250 MIDI files of Classical Piano Music
- Each file was from 0.5 to 10 minutes in length
- Additional set of ~200 simple MIDI piano music files

Outline

- Output Evaluation
 - Indirect Sampling Likelihood (ISL)
- Models
 - 2-Layer LSTM
 - Biaxial LSTM
- Pre-processing / Pre-training
 - Input/Output Reversal
 - Curriculum Learning
 - Key Normalization
- Postprocessing
 - Windowing-RBM

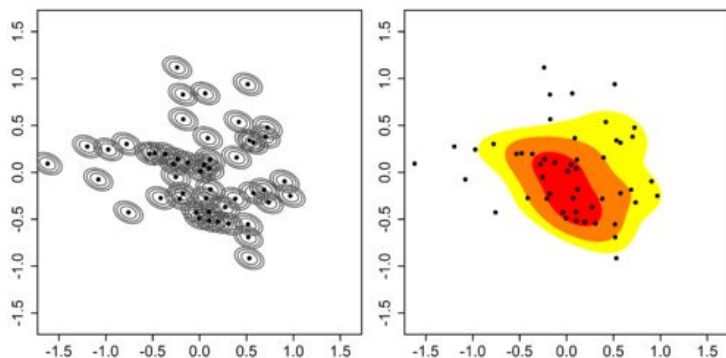
Output Evaluation

In search of an objective measure of generated output...

Indirect Sampling Likelihood:

Estimate the Probability Density Function of a generative model based on its generated output, then compute the log-likelihood of a held-out test set under this PDF.

Compute Density Estimate



Compute Log-Likelihood

$$= f(x | \theta) \rightarrow \ln \mathcal{L}(\theta; x_1, \dots, x_n) = \sum_{i=1}^n \ln f(x_i | \theta),$$
$$= \text{ISL}$$

Closer to 0 is better!

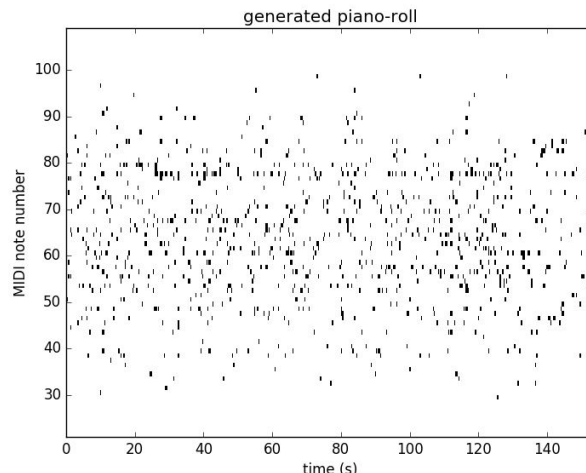
2-layer LSTM

- Architecture

Input	88 Features (piano range)
1st LSTM Layer	512 Nodes
Dropout Layer	Dropout rate: 0.2
2nd LSTM Layer	512 Nodes
Dense Layer	Sigmoid activation

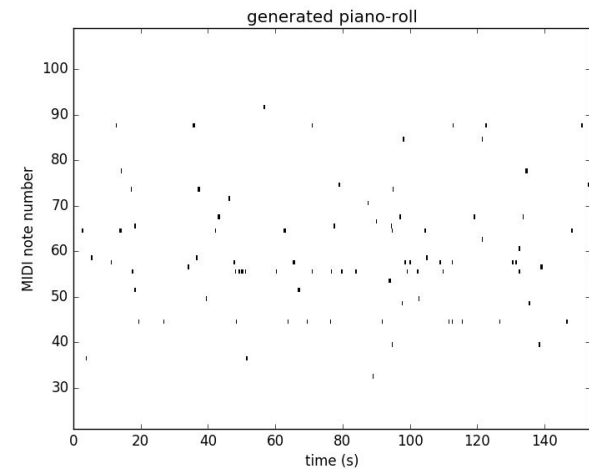
Epoch

10:



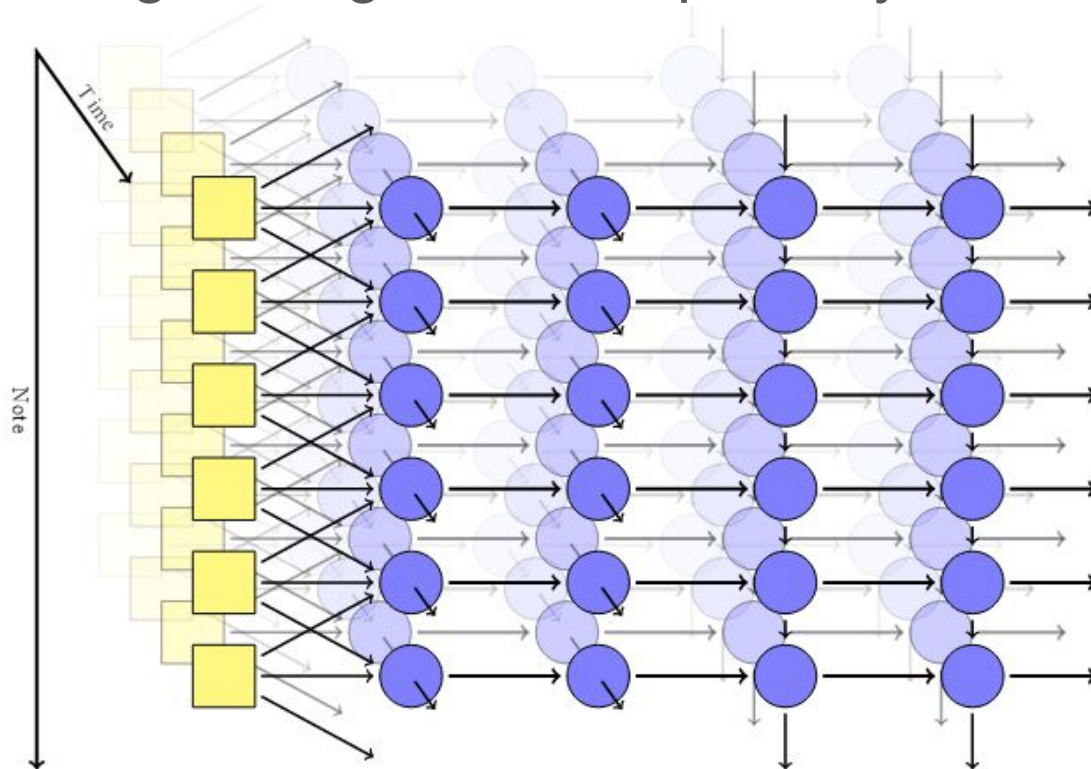
Epoch

70:



Biaxial RNN

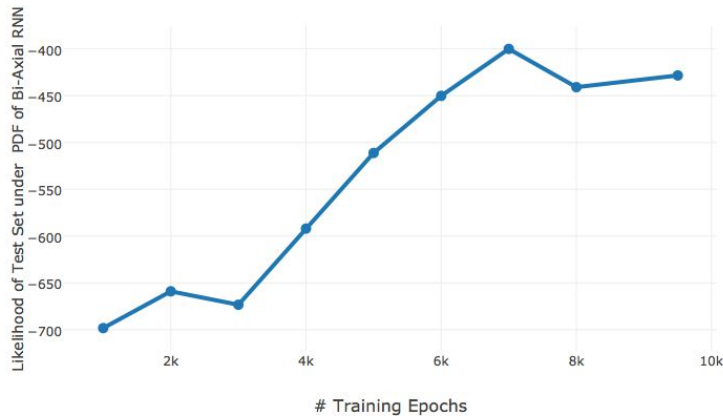
Description: At each unique note, have an LSTM that draws recurrently on itself and also have nodes with lateral connections to neighboring notes. Inspired by CNNs.



Biaxial Results

- ISL vs. Iterations

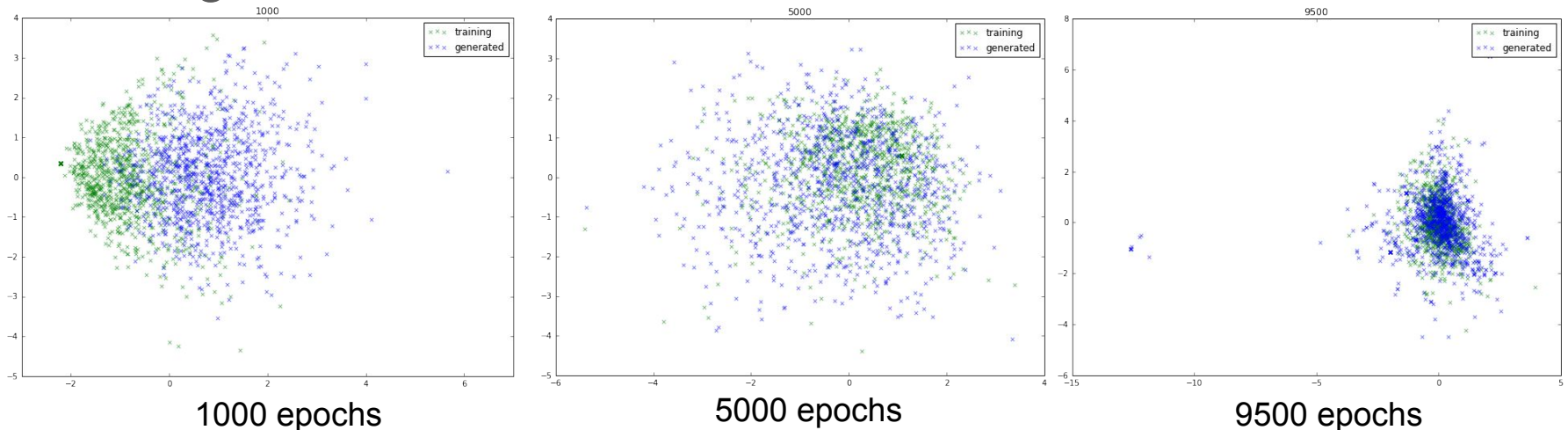
Evaluation of Generative Bi-Axial RNN using Parzen-Window Density Estimation



The ISL measure goes up over training iterations!

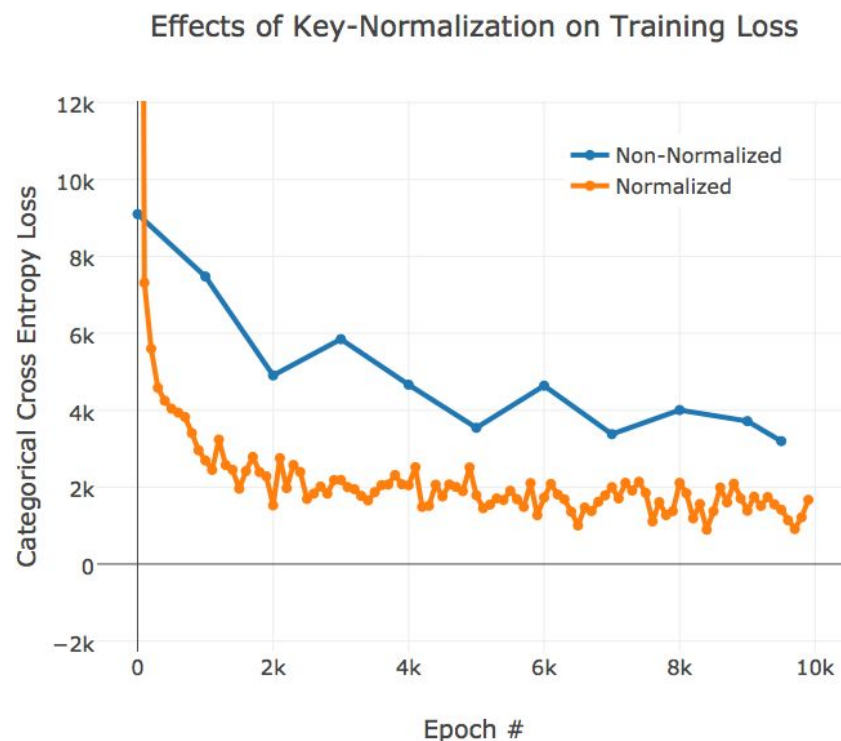
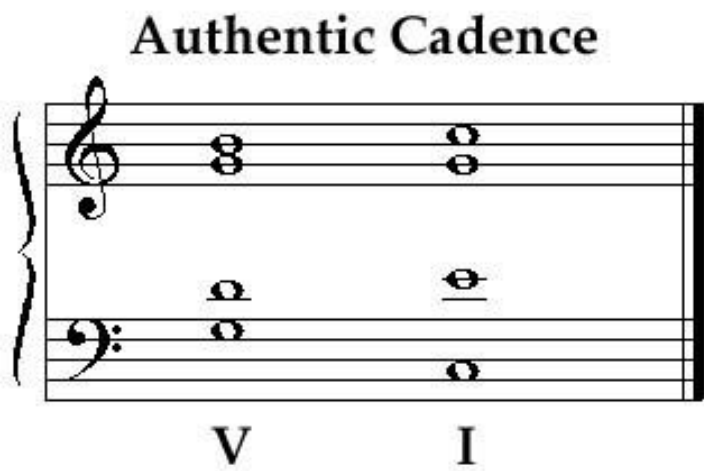
(Remember, the closer to 0 the better, and the Y-Axis is from -700 to -400)

- PCA of vectorized MIDI music (by measure) from training and generated sets



Pre-processing

- Reversing the time axis before training and reversing the output increased train loss, but led to better closure on pieces.
- Key Normalization improved both subjective results and training loss.

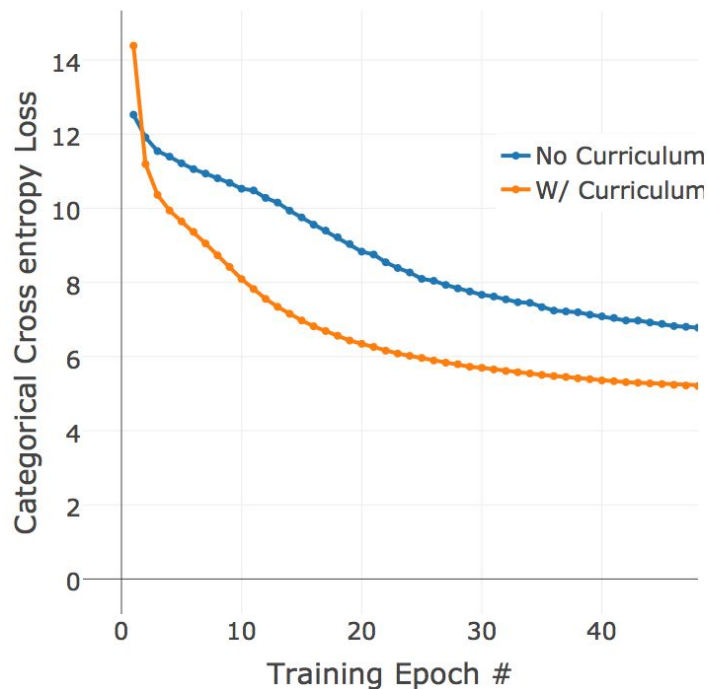


Curriculum Learning

- Pre-train the network on simpler music
- The network trains faster and reaches a lower local minimum

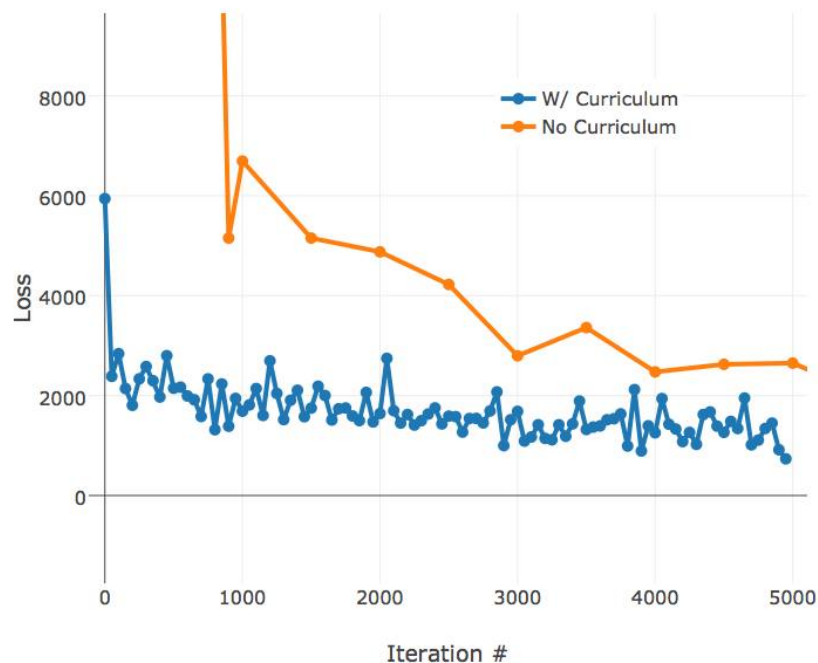
LSTM:

Effect of Curriculum Training on LSTM



Biaxial:

Curriculum Learning vs. No Curriculum on Biaxial RNN



Windowed RBM

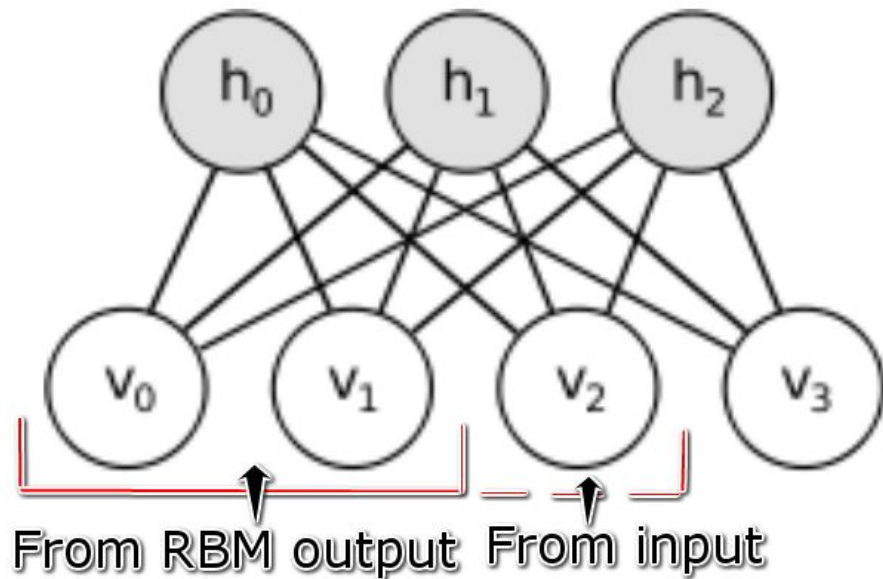
- 2496($16 \times 78 \times 2$) visible units
- 2500 hidden units

Idea

- Map time into space (sliding window of 16 time steps)
- Reconstruct one step at a time
- Reconstruct the whole window at a time

Results

- Always reconstructs the same music piece.
- Even for independent windows.
- Even for empty input.



Conclusions

- There are a lot of difficulties that arise in music generation
- Even measuring output quality is difficult, though similarity to training corpus is a start.
- Classical music seems too complex and varied for current networks to learn any long-term structure.
- Similar principles to other NN tasks, such as data normalization, can make this an easier task.

Questions?

2-layer LSTM Results

- ISL

Epoch	ISL
10	-425.053202876