Viktor Pavlov 230335TAF

1 Problem Statement

In the domain of algorithmic music composition, different strategies can lead to a wide range of outputs in terms of quality and the style of the results. The goal of this study is to explore and compare the musical outputs of two types of models: Music Transformer model and Recurrent Neural Network (RNN) models. The Music Transformer model extends the classic Transformer model with relative attention mechanism, which modulates attention based on how far apart two tokes are. The relative self-attention mechanism allows the model to generate outputs with long-term coherence that extends beyond the length of the training examples. In comparison to the Transformer model, RNNs, including MelodyRNN and PerformanceRNN, are simpler models that can be used to generate monophonic melodies and performance-like polyphonic sequences respectively.

2 Music Transformer

In the course of this study, I utilized two Music Transformer models from Google's Magenta library: Unconditional and Score Conditioned. The Unconditional model, although is capable of generating continuations from a given primer melody (as can be seen in Figure), can also generate music without an initial seed, relying on its understanding of musical patterns and structure. On the other hand, the Score Conditioned model, works in a sequence-to-sequence manner by generating new interpretations of the provided score. This includes capabilities like generating new accompaniment for a given melody, similarly to adding a new dimension to the input (illustrated in Figure). Both models were trained on the MAESTRO dataset. The MAESTRO (MIDI and Audio Edited for Synchronous TRacks and Organization) is a dataset composed of about 200 hours of virtuosic piano performances captured with fine alignment (3 ms) between note labels and audio waveforms. Additionally, the models used the score-to-performance (score2perf) encodings that enrich the MIDI grid-like representations of music with slight variations in timing and velocity in order to achieve a more human-like sound. Keep in mind

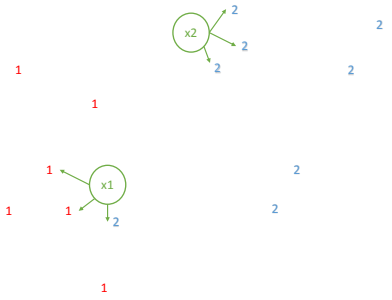

Figure 1: Geometric interpretation of the equidensity contours of the Gaussian.

that all the figures should be referred in text. For example, in Fig. 1 an example from the first lecture is shown.

3 Exercise 3

If you intend to present mathematical equations, please refer them properly in text. For example, the Fisher score is defined by (1).

$$F = \frac{\sum_{j=1}^{C} p_j(\mu - \mu_j)^2}{\sum_{j=1}^{C} p_j \sigma_j^2}. \quad (1)$$

In (1) p_j denotes the fraction of points belonging to the class j. The mean values of the entire data set and the class j are denoted by μ and μ_j, respectively. Finally, σ_j is the standard deviation within the class j. Hint: Do not waste space and time with the equations from the slides.

4 Exercise 4

Table environments can be complicated to learn, but allow one to save a lot of space in reporting the results. Do not forget to

Table 1: Confusion matrix - Decision Tree Classifier

	Actual (PD)	Actual (HC)
Predicted (PD)	5	1
Predicted (HC)	1	4

refer to the Tab. 1 in text.

5 Exercise 5

Please cite all the external sources you have used. As an example, this report is generated in LaTeX[2], which is based on TeX described in[1]. Do not cite lecture slides or the book by C. Aggarwal.

6 Exercise 6

Please follow the two-page limit for home assignment and the 4-page limit for the final project. Be creative and follow the guidelines. Good luck!

References

[1] Donald E. Knuth. *The TeX Book*. Addison-Wesley Professional, 1986.

[2] Leslie Lamport. *LaTeX: a Document Preparation System*. Addison Wesley, Massachusetts, 2 edition, 1994.