CIS 620, Advanced Topics in Deep Learning, Spring 2024

Homework 8
Due: Monday, March 25, 11:59pm
Submit to Gradescope

Learning Objectives

After completing this assignment, you will:

- You will be able to design and implement a neural network to generate audio

Deliverables

This is a **pair** assignment for both the written and coding portions. - **One submission per pair** *(you may also do this individually if you wish)*

1. **A PDF with your name in the agreement**

   Copy and edit this googledoc to enter your names in the Student Agreement and answer the questions mentioned below. **Don't forget to add your team member on the gradescope submission!**

2. **hw5.ipynb file with the functional code**

   Complete the coding assignment in the Jupyter Notebook and upload the file to Gradescope. **Don't forget to add your team member on the gradescope submission!**

Note that there is a separate assignment for the papers, which will be handed in separately.

Note that there is a separate assignment for the project proposal slides!
This is to be submitted (one per project team) via a google form.

Homework Submission Instructions

Written Homeworks

All written homework must be submitted as a PDF to Gradescope. **Handwritten assignments (scanned or otherwise) will not be accepted.**

Coding Homeworks

All coding assignments will be done in Jupyter Notebooks. We will provide a .ipynb template for each assignment as well as function stubs for you to implement. You are free to use your own installation of Jupyter for the assignments, or Google Colab, which provides a Jupyter Environment connected to Google Drive along with a hosted runtime containing a CPU and GPU.

Questions

1. **Implement a neural network that generates music and answer the following questions (see resources below for suggestions on model and dataset).**
    a. **What is the design for your approach?**
    We used a single LSTM model to process the embedded note tokens.

    b. **What is the input data and what dataset was used?**
    We are using the pre-processed music notes data from the GitHub repository provided in the resources link 1. The original dataset is from the resources link 2.

    c. **What is the metric used to evaluate the generated music?**
    We use cross-entropy loss to train the model.
    We use the accuracy to evaluate our model based on the test set.

2. **What are the drawbacks of the current metrics used for understanding the quality of generated music?**
    It seems like an autoregression model, the model won't try to predict the same as the labels provided in the test set, which results in a low accuracy but we obtain similar human evaluations.

    Also, we cannot use deterministic decoding for our model, because it consistently outputs the same index(note) as we decode. To make use of the output distribution, we do a sampling on the distribution, which means other probabilities are also useful in this case compared to just using argmax.

3. **What are the limitations of your current approach? How would you address it?**
    Our approach makes it hard to remember the long-term pattern of the notes, which causes bad music continuity. We might use a transformer-based model in the future to capture the global pattern of the music.

Resources:

1. https://cs230.stanford.edu/projects_fall_2019/reports/26258004.pdf
2. https://magenta.tensorflow.org/datasets/maestro

Class presentation slides

4. (Do not submit as part of this document, submit to google form) Each team will be given 8 minutes to present in class on Thursday, followed by 5 minutes of discussion. Please submit any presentation slides you want to use. You should cover.
    a. Problem Statement/Motivation
    b. Baselines
    c. Metrics
    d. Dataset
    e. Prior Work