

Homework 1

An Introduction to Neural Networks

11-785: INTRODUCTION TO DEEP LEARNING (FALL 2020)

OUT: **September 6, 2020**

Early Deadline/HW1P2 MCQ Deadline: **September 16, 2020, 11:59 PM**

DUE: **September 27, 2020, 11:59 PM**

Start Here

- **Collaboration policy:**

- You are expected to comply with the [University Policy on Academic Integrity and Plagiarism](#).
- You are not allowed to talk work with other students on homework assignments
- You can share ideas but not code, you must submit your own code. All submitted code will be compared against all code submitted this semester and in previous semesters using MOSS.

- **Overview:**

- **Part 2:** This section of the homework is an open ended competition hosted on Kaggle.com, a popular service for hosting predictive modeling and data analytics competitions. The competition page can be found [here](#).
- **Part 2 Multiple Choice Questions:** We also require you to complete multiple choice questions regarding the data. It is recommended you complete the quiz prior to creating and running your models in Homework 1 Part 2. We recommend this as most of you are not familiar with these datasets. Additionally, it saves us time in-office hours. Understanding the data saves you sufficient time with the data loader and allows you to focus on the model itself.

- **Submission:**

- **Part 2:** See the the competition page for details.

1 Part 2: Frame Level Classification of Speech

This part of the homework is a live competition on [kaggle](#).

In this challenge you will take your knowledge of feedforward neural networks and apply it to a more useful task than recognizing handwritten digits: speech recognition. You are provided a dataset of audio recordings (utterances) and their phoneme state (subphoneme) labels. The data comes from LibriSpeech corpus which is derived from audiobooks that are part of the LibriVox project, and contains 1000 hours of speech sampled at 16 kHz. If you have not encountered speech data before or have not heard of phonemes or spectrograms, we will clarify the problem further. For more information, see the paper "LibriSpeech: an ASR corpus based on public domain audio books", Vassil Panayotov, Guoguo Chen, Daniel Povey and Sanjeev Khudanpur, ICASSP 2015 (submitted) ([pdf](#)).

You will be evaluated on the accuracy of the prediction of the phoneme state labels for each frame in the test set.

Please refer to the [kaggle page](#) for more details on the task.

Data

- **train.npy**: (22002,)
- **train_labels.npy**: (22002,)
- **dev.npy**: (2332,)
- **dev_labels.npy**: (2332,)
- **test.npy**: (2251,)

The audio data has been transcribed into mel spectrograms. You have 100 13-dimensional mel spectral (row) vectors per second of speech in the recording.

1.1 Task

Your task is to generate predictions for the phonemes of the test set. You will be evaluated based on the accuracy of your predictions. Grade cut-offs are released after the early deadline. For detailed information, please look at the [kaggle page](#)