

Spoken Digits Recognition

Daniele Gilio

May 15, 2020

0.1 Introduction

This assignment is about spoken digits recognition. The dataset is comprised of 1760 training samples, 120 validation samples and 120 test samples. The dataset has already been processed so that each sample represents the spectrogram of a given digit. Spectrograms are visual representation of audio, the power of the recorded sound wave is divided between 16 frequencies and 64 time periods creating images like the one in Figure 1. Our objective is to create a Neural Network (a Multi-Layer Perceptron to be precise) in order to recognize the spoken digits.

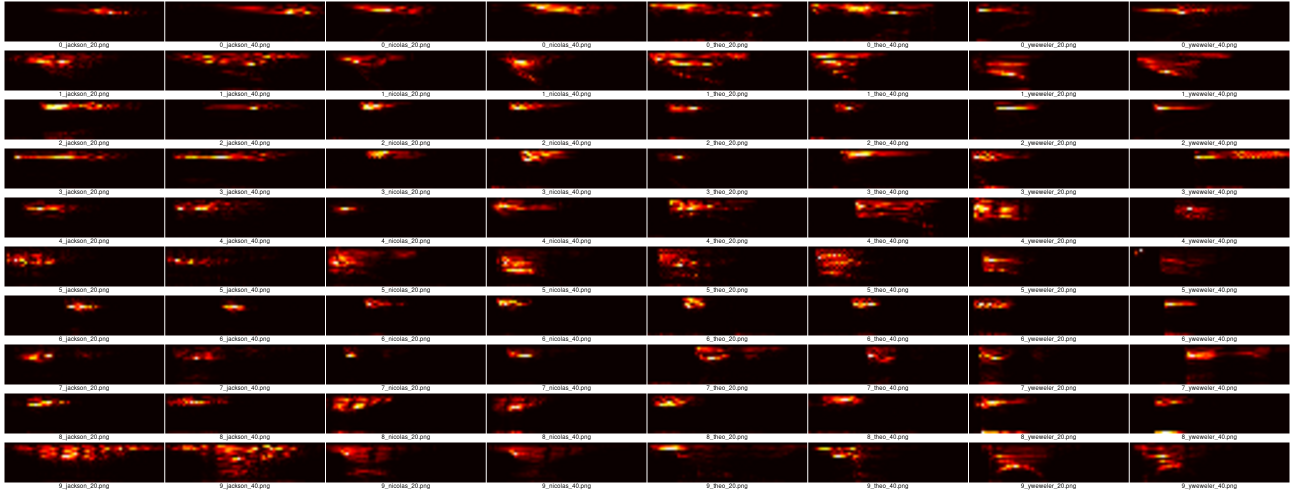


Figure 1: Example of the dataset spectrograms

0.2 Data Visualization and Analysis

In order to test our code we plotted spectrograms after we loaded the data and after performing normalization. The second plotting was also done in order to verify that the normalization did not interfere too much with the data. At this stage we opted for an L2 Normalization since we feel that, since the data in a single vector are probably strongly correlated, avoiding statistic normalization might be a feasible choice. We can see the output of our code in Figure ().

0.3 Neural Network

In order to experiment a bit with Neural Networks we built a Single-Layer Perceptron in which we have only an input layer and an output layer. As in the previous assignment we implemented a varying learning rate.

0.3.1 Architecture

After playing a bit with network architectures we settled on the following sequence [1024, 256, 128, 32, 10].

0.3.2 Hyperparameters

0.3.3 Normalization Techniques

0.4 Results

0.5 Conclusions