Homework # 3: Portion # 3

Due Date: Monday, October 28, 2019 by 11:59PM

Total Points: 65

This assignment provides practical programming experience with training neural networks for speech enhancement. **For the purposes of this assignment, please work individually**. However, feel free to ask basic questions to the instructor and classmates, but complete the assignment on your own. Feel free to search for information (inputs and outputs) about key Python functions, using the internet. From Canvas, go to Files and download all content from the HW#3_partIII folder. You should also use code from the earlier parts of HW#3.

Submit your answers to the questions and Python code to Canvas. **Be sure to include a README that explains each major script/function.** Feel free to create separate functions as needed. **Be sure to comment your code and to submit all files to Canvas**.

All assignments must be submitted on time to receive credit. No late work will be accepted, unless you have a prior arrangement with the instructor.

Question 1. [15 POINTS]

Data Loading and Preprocessing. Follow the Data Pre-processing tutorial (e.g. datapreprocessing.ipnyb) to create data loading objects for the training, development and testing sets. Note that there are comments in the code that need to be addressed. In addition to this, perform the steps below. **Note that this tutorial serves as a guide. Feel free to modify based on your comfortability and experience. Also note that the general framework should be kept the same (e.g. single time-frame serves as input and output.)**

1. For the training and development datasets, modify the code to compute the means and standard deviations across all time frames and for each frequency. Plot (in separate subplots) the resulting mean and standard deviations for both sets versus frequency (in Hz). Be sure to correctly label the plots.

2. Likewise, compute the minimum and maximum values for each frequency, across all time frames of the training and development datasets (e.g. one min and max vector for each set). Plot (in separate subplots) the resulting min and max vectors for both datasets versus frequency (in Hz). Be sure to correctly label the plots.

3. Using the mean and standard deviations, perform data standardization on the training and development datasets. After this is performed, recompute and plot the mean and standard deviation of the standardized data for both datasets.

4. Using the minimum and maximum vectors, perform data normalization on the training and development datasets. After this is performed, recompute and plot the minimum and maximum values of the normalized data for both datasets.

Question 2. [50 POINTS]

DNN For speech enhancement. In this problem, you will train separate DNNs that each estimate their own training target. Use your data from the prior homework assignment and the data loader from the prior problem to train the networks. **Separately run the experiments below for the data that is modified by normalization and standardization approaches**. Specific instructions follow below.

1. Write separate functions to compute the fast Fourier Transform (FFT) mask, ideal binary mask (IBM) and ideal ratio mask (IRM) given the speech and noise segments of a noisy speech signal. For the FFT-mask, truncate the label to values between 0 and 1 (inclusively). Take one of your noisy speech signals (and corresponding speech and noise components) and generate the IBM and IRM masks. Plot the spectrograms for the noisy speech signal, clean speech component, noise component and appropriately label all axis. Also, generate plots for the corresponding IBM and IRM.

2. Using the above functions, train separate DNNs that individually estimate the clean speech spectrogram, FFT mask, IBM, and IRM. In other words, you should have four different DNNs, one for each training target. Use the parameters and network structure that is outlined in the paper. Note that example DNN code is included in the homework folder. Be sure to test the network with the validation/development data after each Epoch and perform model selection with the best performing result. For each DNN, generate and plot error curves (MSE) as a function of Epoch for the training and validation/development set. Discuss how the different DNNs perform. Also, discuss how data normalization vs data standardization impacts performance.

3. After the networks are trained, test each of the networks with the testing data set. Generate plots that show the average MSE between the estimated and true clean speech time-domain signals for each training target. Also, listen to the signals and explain how the performance varies audibly. Also submit two sound examples (clean speech, noisy speech, enhanced speech) for each training target. Describe how the different targets perform at speech enhancement, both computationally and perceptually.