Deep Learning Course (Assignments)

General comments

The practical part of this course revolves around a speech classification task. For assignment programming, we recommend you to install PyTorch (https://pytorch.org/) with Anaconda (https://www.anaconda.com/distribution/). The following guide might be of help: https://deeplizard.com/learn/video/UWlFM0R_x6I.

Your job is to develop different neural network architectures that classify the words "yes", "no" and *other word* (3-class classification problem) from speech recordings. You are provided with labeled (log-Mel) speech feature matrices extracted from one-second long speech segments comprising one word each that can be downloaded from

https://www.dropbox.com/s/ed6kycb0bc65q56/DL_Data.rar?dl=0

The above compressed file contains the following pickle Python files:
X_train.p – 9,489 speech feature matrices for training
Y_train.p – Ground truth labels for *X_train.p* in one-hot encoding
X_valid.p – 1,198 speech feature matrices for model validation
Y_valid.p – Ground truth labels for *X_valid.p* in one-hot encoding
X_test.p – 1,229 speech feature matrices for testing your models
Y_test.p – Ground truth labels for *X_test.p* in one-hot encoding

In the next table, you can check how many samples of the different classes you have available to work with and how they are distributed per data set:

Class / Data set	Training	Validation	Test
"yes"	3,228	397	419
"no"	3,130	406	405
other word	3,131	395	405
TOTAL	9,489	1,198	1,229

The dimension of each speech feature matrix is 101x40, where 101 is the number of time frames and 40 is the number of frequency bins. Below you can see an example of how the word "yes" looks like. Enjoy solving the assignments!