Machine Learning for IOT - Homework 2

Sepehr Alidousti Shahraki, Shadi Nikneshan, Ali Ghasemi Group 22 January 2022

Exercise 1

In this exercise we first create our own window generator function in order to create the dataset we want to feed to our model (more information about the code is available in the comments in the code file). We tested three different models for two scenarios in the homework file. We test LSTM, MLP and, CNN (LSTM model didn't give us good enough results so we avoided using it in the report). For both output steps of 3 and 9, we used the CNN model. Before the fitting process, normalization was performed on the data. After the fitting process, the pruning procedure was performed in order to reduce the size of the TFLITE model file as much as possible. We have reduced the number of units using the value alpha in order to reduce the complexity of our model. The best value available for alpha was 0.25. This value is obtained after experimenting with different values in the range of 0.25 to 0.75. The values for metrics and loss in the fitting process are as mentioned in the lab sessions and for the pruning process the values for initial sparsity and final sparsity have been found after experimenting with different possible values. You can find the results of the models in the table below. The MAE values are measured after the pruning process; The selected models are **highlighted** in the table.

Model	MAE(Temperature)	MAE(Humidity)	TFLITE size(KB)	Output Steps
MLP	0.028665833	0.06857006	1.85	3
MLP	0.13139595	0.05768358	2.2	9
CNN	0.033870477	0.073597886	1.6	3
CNN	0.24555247	0.1395972	1.7	9

Exercise 2

In this exercise we first chose different values for frame length, frame step, number of Mel bins, frame steps and, MFCC (using it or not using it) in order to get the best combination for the data that we're using for our model. After this step different ML and Deep Learning models (DS-CNN, CNN, MLP) had to be tested in order to find the best model and the best hyperparameters for them. The best combinations of models and their hyperparameters and the best combination (that satisfy the requirements for the problem) for the parameters that create our data and their results (accuracy and latency), and their sizes are mentioned in the table below. The value for alpha is the value that we use in order to control the number of units in each layer of our CNN model. We used magnitude-based pruning techniques in order to reduce the size of our TFLITE model and after that, compression was performed to reduce the file size even more.

Model	Problem number	Alpha	Momentum	Epochs	Frame Length	Frame Steps	Number of Mel Bins	Number of Coefficients	Number of Filters	Accuracy	Size (KB)	Latency (ms)
CNN	a	1	0.1	30	650	350	30	9	512	0.935	3000	-
CNN	b	1	0.3	30	650	350	30	9	256	0.933	956	68.28
CNN	c	1	0.6	30	650	350	30	9	128	0.92	246	57.81

Command for latency: python kws_latency.py --mfcc --length 650 --stride 350 --bins 30 --coeff 9