Machine Learning for IoT - Homework 01

Francesco Di Salvo s282418 | Francesco Lacriola s292129 | Gianluca La Malfa s290187

Exercise 1 - Temperature and Humidity Dataset with TFRecord

The goal of the exercise was to build a TFRecord Dataset, starting from a CSV file containing the recordings gathered through the DHT11 sensor. Considering that the DHT11 sensor has an 8 bit resolution (i.e. Integer), we used an Integer representation. Here, we obtained an output size of 1722 B.

Then, we used a min-max normalization, considering the minimum and the maximum values reported in the technical datasheet, that are $[0, 50]$ and $[20, 90]$ for the temperature and humidity, respectively. Since the normalization process returns decimal numbers between 0 and 1, in order to preserve the information, we used a Floating Point representation, obtaining an output size of 1848 B.

Exercise 2 - Audio pre-processing optimization

The goal of the exercise was to improve the execution time of the proposed preprocessing pipeline for the extraction of the The Mel Frequency Cepstral Coefficients (MFCCs). In order to compute it, we need the mel spectrogram, that is the product between the spectrogram and the *linear to mel weight matrix*.

This matrix is defined through many parameters, but the only "non-fixed" one is the *number of spectrogram bins*, which is defined as the half of the frame length plus one. It is the "half" because the spectrum of a real signal has an Hermitian symmetry, hence, roughly half of the spectrum bins are redundant. Therefore, we can define the matrix in advance just once, rather than computing it for all the spectrograms.

Another way for improving the execution time was to *reduce the sampling rate*. We performed further tests with a sampling rate of 8 kHz, but in order to obtain the same number of frames of the proposed MFCCs, we needed to adjust both *frame length* and *frame step* values, according to the equation:

$$\text{floor}\left(\frac{\text{Rate - F. length}}{\text{F. step}}\right) + 1 = 124$$

where 124 is obtained on the slow preprocessing pipeline. Moreover, by using 16 mel bins, we were able to reduce the execution time up to 14.48 ms, while preserving a sufficient Signal to Noise ratio (18.21 dB). As concerns the number of coefficients, we needed to set them to 10 to be in line with the shape constraint. A better signal reconstruction (56.20 dB) can be obtained with 32 mel bins, obviously at the expense of the execution time.

Finally, we did not modify the lower and upper frequencies, because they do not affect the execution time.

Table 1: Preprocessing benchmarks

Rate	Frame length	Frame step	N. Mel Bins	ET	SNR
16 kHz	16	8	16	16.65 ms	13.73 dB
8 kHz	8	4	16	**14.48 ms**	18.21 dB
8 kHz	8	4	32	15.04 ms	**56.20 dB**