CSE 421 – Embedded Machine Learning
Homework 4 – Q2 (Section 11.7): Keyword Spotting

Can Sarp Aridogan
Ahmet Renda Sendag

1. Objective

This task implements a keyword spotting system using audio signals. The goal is to distinguish the digit zero from all other spoken digits using a lightweight model suitable for embedded systems.

2. Dataset and Feature Extraction

The dataset consists of recorded spoken digits stored as WAV files. Each audio signal is resampled to 8 kHz and processed using Mel-Frequency Cepstral Coefficients (MFCC). For each recording, 13 MFCC coefficients are extracted per frame, and the mean and standard deviation of each coefficient are concatenated to form a 26-dimensional feature vector.

3. Model

A single-neuron logistic regression model with sigmoid activation is trained on the MFCC-based features. Binary cross-entropy loss and the Adam optimizer are used. Class weighting is applied to reduce false negatives for the zero digit.

4. Results

The trained model achieved an accuracy of 0.88 on the test set. The number of false negatives (zero classified as not zero) is 38. The confusion matrix (rows = true label, columns = predicted label) is given below.

[12 38]
[22 428]

5. Embedded Implementation Plan (Mbed/STM32)

After training, the learned weights and bias are exported to C source code. On the STM32 platform, MFCC features are computed from incoming audio frames, and the logistic regression model is evaluated using a dot product followed by a sigmoid function. The output is thresholded at 0.5 to produce the final keyword decision.

6. Conclusion

The results show that MFCC features combined with a very small neural model can effectively perform keyword spotting. The approach is computationally efficient and well suited for deployment on embedded microcontrollers.