电子科技大学
格拉斯哥学院
Glasgow College, UESTC

UESTC 3036: Machine Learning & AI

Fall Semester 2023-2024

Lab 1 - Week 4

Spoken Digit Recognition

Report Due Date: Friday, 29 September 2023 @ 23:59 CST

Student's Chinese Name	
Student's English Name	
Student's UESTC ID	
Student's UoG ID	

Pre-processing, Feature Extraction, and Classification

1. Overview

In this lab, you will complete a spoken digit recognition using a classic machine learning algorithm, K-Nearest Neighbors (KNN) and Support Vector Machine (SVM). The task is based on a small audio dataset called *Audio-MNIST*.

2. Dataset

Audio MNIST is a simple audio/speech dataset consists of recorded spoken digits in wave files sampled at 8kHz. The recordings are trimmed to have near-minimal silence at the beginning and end.

The dataset contains 3000 spoken digits. For each sample, the input is an audio wave file, and the output should be an integer between 0-9. The raw audio is put in `audio_mnist/recorddings`, where files are named in the following format: `{digitLabel}_{speakerName}_{index}`.wav. For Example, `7_jackson_32`.wav. The dataset is available as part of the "*Lab1_Code-and-Dataset.zip*" folder.

3. Scikit-Learn Package (official documentation)

Scikit-learn, also called sklearn, is an open-source Python-based Machine Learning toolkit. It implements efficient algorithm applications through Python numerical computing libraries such as `NumPy`, `SciPy`, and `Matplotlib` and covers almost all mainstream Machine Learning algorithms.

To install sklearn, just use the pip command in your command line.

```
pip install -U scikit-learn
```

4. Lab Procedure

1. Read the raw data and play it in Jupyter notebook.
2. Read the full name of each data file into a Python list.
3. Split the given dataset into training and test sets. The dataset contains 3000 samples. You will split it into a training set (2100 samples) and a testing set (900 samples), i.e. 70% to train the models and 30% to evaluate the performance. The testing set, which the model has never seen, is used for measuring its performance.
4. Read raw data from the audio file and stack them as a tensor (matrix). Each audio stream will be read as a numpy array. We then extract the MFCC as its features, which convert each data into a vector (20 MFCC coefficients from each frame and an overall array with 120 dimensions). By stacking the 2100 and 900 vectors, the training set and the testing set will be a tensor (in this 2-dimensional case, it is a matrix) with the dimension of (2100, 120) and (900, 120).

5. Apply the data standardisation method to normalise the data using the formula given in (1). Then, plot the standardised feature of the first data sample in the training set.

$$\boldsymbol{z} = \frac{\boldsymbol{x} - \boldsymbol{\mu}}{\boldsymbol{\sigma}} \tag{1}$$

   where:
   x: raw data
   μ: the mean
   σ: the standard deviation

6. Implement K-Nearest Neighbors algorithm and present the classification performance with the confusion matrix and F1-score. In the given implementation, we used `n_neighbors` = 5.
7. Implement a Support Vector Machine and present the classification performance with the confusion matrix and F1 scores.

5. Lab Tasks and Report

- Explain each step (1 to 7) of the lab procedure and its output in detail. In addition, you are required to critique the results obtained at each stage of the lab procedure. [40 marks]
- Keeping the KNN and SVM parameters fixed as given in the lab procedure, try different numbers of MFCC features i.e., n_mfcc = 10, 20 and 30, and present the classification performance in a table format. Compare and comment on the classification performance. [20 marks]
- Keeping the MFCC features, i.e., n_mfcc = 20, evaluate the performance of the KNN algorithm with different numbers of neighbors i.e., n_neighbors = 3, 5, and 7. Compare and comment on the classification performance. [20 marks]
- Keeping the MFCC features i.e., n_mfcc = 20, evaluate the performance of the SVM with three different kernel functions. Compare and comment on the classification performance. [20 marks]