CENTER FOR SPEECH AND LANGUAGE TECHNOLOGIES

LUMS

Project Phase II: Gender & Speaker Recognition

Instructions:

- The aim of this project is to give you a hands-on with a real-life machine learning application.
- Use separate training, validation and testing data as discussed in class.
- This is a group project and maximum number of members per group is 5. **Only one submission per group is required.**
- **Carefully read the submission instructions, plagiarism and late days policy at the end of project.**
- Deadline to submit phase II is: **Wednesday, 2nd December 2020.**

Problem:

The purpose of this project is to get you familiar with Multinomial Logistic Regression (aka Softmax Regression) for Gender Classification and Speaker Recognition using speech. Gender Recognition is the task of classifying gender (using binary gender classification: male/female) of speaker using their voice. Speaker Recognition problem is defined in many ways but the one you will be implementing is a closed set speaker recognition i.e. given N speakers classify the test utterance (recording) into one of the N speaker classes. You are given a speech dataset that contains 1,420 recordings, tagged with gender and speaker information. Your task is to train a Multinomial Logistic Regression classifier and report evaluation metrics on the test set.

Dataset:

The core dataset contains 1,420 recordings from 142 (109 males and 33 females) speakers (10 recordings per speaker). The details of the dataset-split for both problems are as follows:

Speaker Recognition:

Train on a dataset D of N speakers, and then classify a test utterance x into one of N classes. There is no overlap of utterance between the training and test sets. However, there is complete overlap between training and test speakers. The dataset is divided utterance (recording)-wise into three parts:

- Train: 852 recordings (6 recordings per speaker)
- Valid: 284 recordings (2 recordings per speaker)
- Test: 284 recordings (2 recordings per speaker)

Gender Recognition

Train on a dataset D of male and female genders. Classify a test utterance x as male or female. There is no overlap of speaker or utterances between the training and test sets, in order to prevent speaker characteristics from helping with gender classification. The dataset is divided speaker wise into three parts:

- Train: 108 speakers (83 male, 25 female)
- Valid: 17 speakers (13 male, 4 female)
- Test: 17 speakers (13 male, 4 female)

In both splits there are three top-level directories [train/, valid/, test/] corresponding to the training, validation and testing sets. Each contains directories [e.g. SPK001_M] for the speaker recordings where SPK001 is speaker ID and M is gender. For speaker recognition your label would be speaker ID (e.g. SPK001) whereas for gender recognition your label would be gender (e.g. M for male). Within these directories, recordings are stored in WAVE format.

You can download the dataset from this link.

Feature Extraction:

In the feature extraction step you will represent each WAVE file by 13-dimentional Mel-frequency Cepstral Coefficients (MFCCs). MFCCs are a widely used representation for human speech. A code snippet is provided here that shows how to install the required library and read & represent a sample WAVE file with a MFCC vector.

Note: You yourself have to append x_0 = 1 to handle bias.

Implementation:

Implement Multinomial Logistic Regression from scratch keeping in view all the discussions from the class lectures. Feel free to read Chapter 5 of Speech and Language Processing book to get in-depth insight of Multinomial Logistic Regression classifier. Specifically, you'll need to implement the following:

- Softmax function
- Cross-entropy loss function (for multinomial logistic regression)
- Batch Gradient Descent
- Prediction function that predict the label of test recordings using learned multinomial logistic regression
- Evaluation function that calculates classification accuracy, macro-average (precision, recall, and F1) and confusion matrix on test set.
- Report plots with no. of iterations/ epochs on x-axis and training & validation loss on y-axis. Try out different combinations of learning rate and epochs.

Notes:

- You need to implement Multinomial Logistic Regression once but you need to run it twice (independently) for two different problems (1) Gender Recognition (2 output classes) and (2) Speaker Recognition (142 output classes).

- Use the provided validation set to find the optimal learning rate and epochs.
- You can use scikit-learn's accuracy score function to calculate the accuracy, classification report to calculate macro-average (precision, recall and F1) and confusion matrix function to calculate confusion matrix on test set.
- The expected macro-F1 for Gender Recognition is around 80% and Speaker Recognition is around 93%.

Use the procedural programming style and comment your code thoroughly.

Submission Instructions:

Submit your code both as notebook file (.ipynb) and python script (.py) on LMS. The name of both files should be the roll numbers of group members e.g. rollno1_rollno2.ipynb. If you don't know how to save .ipynb as .py see this. **Failing to submit any one of them will result in the reduction of marks**.

Plagiarism Policy:

The code MUST be done independently. Any plagiarism or cheating of work from others or the internet will be immediately referred to the DC. If you are confused about what constitutes plagiarism, it is YOUR responsibility to consult with the instructor or the TA in a timely manner. No "after the fact" negotiations will be possible. The only way to guarantee that you do not lose marks is "DO NOT LOOK AT ANYONE ELSE'S CODE NOR DISCUSS IT WITH THEM".

Late Days Policy:

The deadline of the project is final. However, in order to accommodate all the 11th hour issues there is a late submission policy i.e. you can submit your project within 3 days after the deadline with 25% deduction each day.