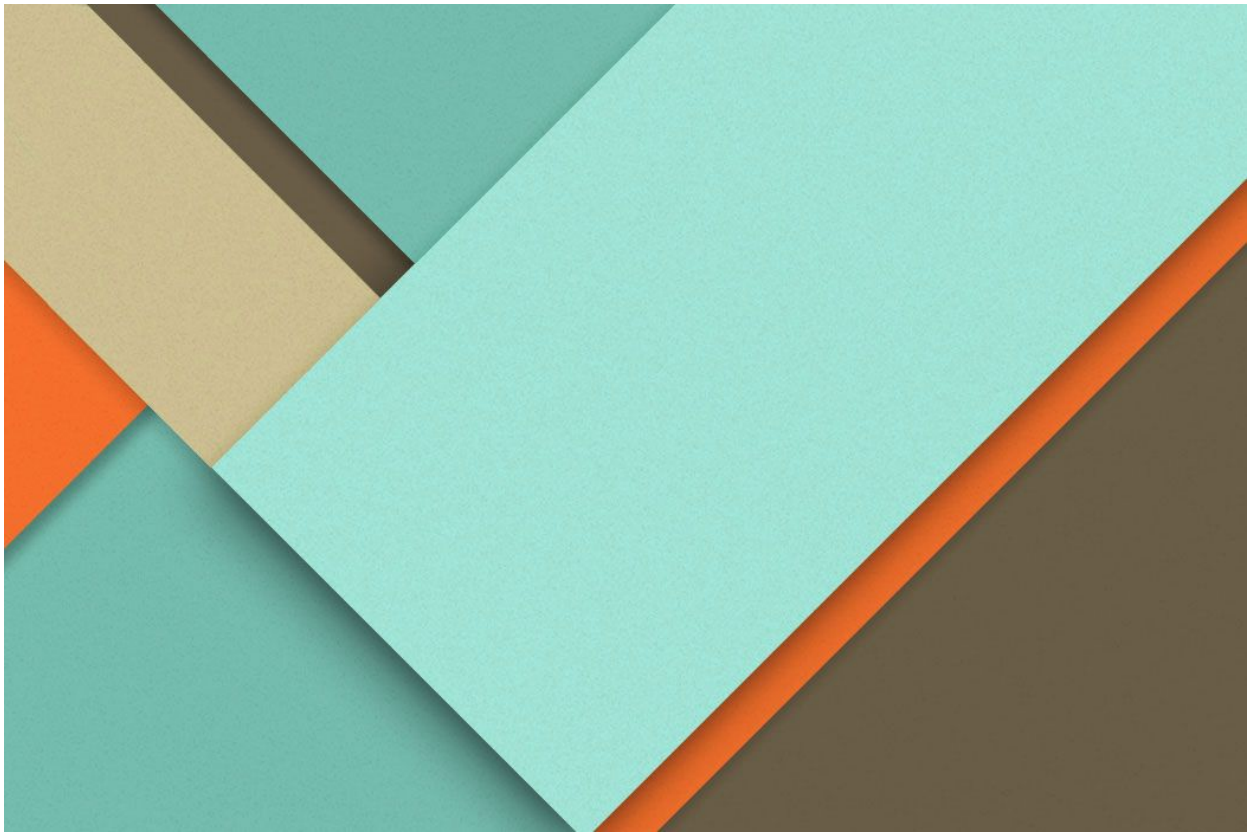

Assignment-2

Dialect classifier

Kartik Gupta IMT2016128
Pavan Kumar Pothula IMT2016127
Lalith Kota IMT2016132
Akash Sharma IMT2016124

Problem Statement

Given a data set of audio files containing speech , classify them into 9 IViE british region dialects using machine learning algorithms.

Approach

Dialects are basically how differently a word is pronounced by two individuals. I.e., which syllable of the word is getting more stress. Or technically which frequency range is getting more energy.Hence we found MFCC as a good choice of feature to distinguish different dialects.

Feature extraction

How do we calculate MFCC and how are we doing it is briefly described here:

FIrst we split the each of the audio signals into small pieces/frames, of 25ms length. Then we window the frames with hamming window. Then we take the DFT of each frame. Then we compute the power spectrum of each frame (using the above DFT).

MFC is the mel-scaled-frequency-axis version of the power cepstrum.This Mel Scaling is done by multiplying the power spectrum with a series of overlapping Triangular Windows, with increasing width.

So, we obtain a window by the end of each multiplication. We then take the sum of all the element (of this particular array or window-output). We now take the log of each element. Then take the DCT of each of the values.

By the end of this, we are left with an whole array of fixed no of values for each frame. Now we have to have a generalised form of putting this as our features because each of the audio signal is of varying size and is spoke at varying speeds. So, a frame, that is of constant length, might not have the same spoken signal or content in the same frame of a different dialect (or audio file, for that matter) So we are taking average of each of the same coefficient of each frame. That would give us a final fixed number of values which would exactly be out features.

Implementation

Getting the data in required format:

To implement the different machine learning algorithms, we need to get the data in numerical values with proper labels. The given data set is in the form of 67 audio files (.wav) stored in each of the 9 folders. The folders represent the 9 dialects of the IViE dataset in which we need to classify the the audio files. So we used python_speech_features library python to get the mfcc coefficients for each audio file in an numpy matrix which we later converted into a pandas dataframe. Also we took we noted down the ID of the folder from which the audio file is coming from.

Preprocessing:

For each audio file, we get 26 values for all the frames in which the audio signal was divided it into. So we took the mean value of each coefficient to standardise the data-set. Then we normalised the whole data set . Since the data set was complete, there was no problem like null values or missing values.Then we shuffled the data points.

Training and testing:

We splitted our data set into train and test data in the ration 70:30. We trained SVM classifier model with linear kernel, logistic regression model and K nearest neighbour with K = 3 and 5.

Improving accuracy:

For improving accuracy we tried different values like mean and standard deviation of the mfcc coefficients from different frames to standardise the data points over of the coef. Out of which mean was giving the best result. Also we normalized the data as the data was having a high variance.

Result

We were able to successfully implement a dialect classifier model with best accuracy of 0.93 using SVM classifier with linear kernel. Logistic regression model gave accuracy of 0.85 over the testing data and k nearest neighbour model giving accuracy of 0.85 with k=3 and 0.83 with k=5.