Assignment 2

MFCC or Mel Frequency Cepstral Coefficients:

1. It is a speech/audio representation method which is very well suited for human audio. It is a logarithmic method wherein particular frequencies in the audio are converted to Mel scale on which further speech recognition tasks are performed. They are suitable for human audio because the Mel scale frequencies are good at representing the frequencies at which humans speak/utter words.

Spectrogram

1. Spectrogram method is used to convert the audio signal into time-frequency domain using 2D fourier transform. .The audio signal is segmented and then transformed. Thus this new frequency time domain signal is rich in information about the frequency variations of the audio signal. This information can be used for classification of the audio signal.

Results

MFCC

Prec 0.760389
Recall 0.7201
f1 0.7392

MFCC_noise

Prec 0.52196421268012396
Recall 0.5093428345209818
f1 0.51555771111636164

Spectrogram

SVM:

```
           precision    recall  f1-score   support

        0       0.95      0.77      0.85       260
        1       0.65      0.78      0.71       230
        2       0.57      0.73      0.64       236
        3       0.73      0.74      0.73       248
        4       0.71      0.65      0.68       280
        5       0.66      0.47      0.55       242
        6       0.86      0.78      0.82       262
        7       0.84      0.74      0.79       263
        8       0.64      0.82      0.72       243
        9       0.68      0.71      0.70       230
```

Prec 0.7300617129879543
Recall 0.7197273456295108
f1 0.7248576966577809

Specttogram_Noise

SVM:

```
          precision    recall  f1-score   support

       0       0.95      0.77      0.85       260
       1       0.65      0.78      0.71       230
       2       0.57      0.73      0.64       236
       3       0.73      0.74      0.73       248
       4       0.71      0.65      0.68       280
       5       0.66      0.47      0.55       242
       6       0.86      0.78      0.82       262
       7       0.84      0.74      0.79       263
       8       0.64      0.82      0.72       243
       9       0.68      0.71      0.70       230
Prec  0.7300617129879543
Recall 0.7197273456295108
f1   0.7248576966577809
```

Analysis:

Mfcc slightly outperformed spectrogram as seen from the results. This could be due to the fact that the audio inputs are from human subjects. And as can be seen from the above written points, mfcc is able to represent the human audio better than spectrogram.

However Mfcc fails when I add noise to the mix. Spectrogram works better here. This could be due to the fact that the noise added has no human component to it, thus spectrogram works better on it and mfcc fails to extract meaningful data out of it.

References :

1. Spectrogram
    a. Discussed the code with Priyanshi Jain 2017358
    b. https://towardsdatascience.com/understanding-audio-data-fourier-transform-fft-spectrogram-and-speech-recognition-a4072d228520
    c. https://fairyonice.github.io/implement-the-spectrogram-from-scratch-in-python.html
2. MFCC
    a. https://github.com/jameslyons/python_speech_features/blob/9a2d76c6336d969d51ad3aa0d129b99297dcf55e/python_speech_features/base.py#L149
    b. https://haythamfayek.com/2016/04/21/speech-processing-for-machine-learning.html

c. Discussed the code with Priyanshi Jain 2017358