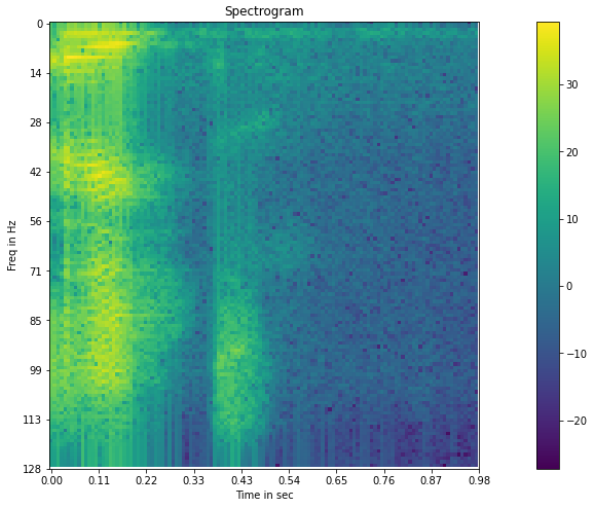
Sejal Bhalla
2017100

MCA Assignment - 2

NOTE: The code for Q3**,** i.e training a SVM on the extracted features is included in the files for Q1 and Q2.

Performance of Spectrogram



The performance of spectrograms was evaluated by calculating the following features and fitting a SVM on it:

1. Spectral Centroid
2. Spectral Crest Factor
3. Spectral Decrease
4. Spectral Flatness
5. Spectral Kurtosis
6. Spectral RollOff
7. Spectral Slope
8. Spectral Spread

By using GridSearch, the following hyperparameters were found to be the best fit for the given data: `{'C': 100, 'gamma': 1, 'kernel': 'rbf'}`

Classification Report:

1. Without Adding Noise:

   Average Precision: **16%**

Sejal Bhalla
2017100

Average Recall: **10%**
Average F1 Score: **8%**
Average Accuracy: **10%**

	precision	recall	f1-score	support
0	0.08	0.45	0.13	243
1	0.09	0.14	0.11	242
2	0.00	0.00	0.00	280
3	0.02	0.02	0.02	230
4	0.08	0.09	0.09	230
5	0.14	0.01	0.01	263
6	0.97	0.29	0.45	262
7	0.17	0.01	0.02	248
8	0.02	0.00	0.01	236
9	0.00	0.00	0.00	260
accuracy			0.10	2494
macro avg	0.16	0.10	0.08	2494
weighted avg	0.16	0.10	0.08	2494

2. With Noise: Noise samples were selected randomly and added after multiplying with a factor = 0.005 to each audio file, adding up to a total of 20000 files for training.

Average Precision: **9%**
Average Recall: **10%**
Average F1 Score: **8%**
Average Accuracy: **10%**

	precision	recall	f1-score	support
0	0.00	0.00	0.00	260
1	0.05	0.06	0.05	230
2	0.10	0.27	0.14	236
3	0.04	0.00	0.01	248
4	0.09	0.02	0.03	280
5	0.11	0.03	0.05	242
6	0.30	0.36	0.33	262
7	0.09	0.01	0.02	263
8	0.07	0.26	0.11	243
9	0.04	0.03	0.03	230
accuracy			0.10	2494
macro avg	0.09	0.10	0.08	2494
weighted avg	0.09	0.10	0.08	2494

Sejal Bhalla
2017100

Performance of MFCCs

The first 12 MFCCs were used for the classification. For each MFCC, the following features were extracted:

1. Mean
2. Median
3. Maximum
4. Minimum
5. Skew
6. Standard Deviation

Total features: 40 MFCCs * 5 features = 200 features
(Mean was excluded since MFCCs were mean normalised)

By using GridSearch, the following hyperparameters were found to be the best fit for the given data: `{'C': 100, 'gamma': 0.01, 'kernel': 'rbf'}`

Classification Report:

1. Without adding noise:

   Average Precision: **64%**
   Average Recall: **61%**
   Average F1 Score: **60%**
   Average Accuracy: **61%**

	precision	recall	f1-score	support
0	0.59	0.78	0.67	243
1	0.40	0.56	0.47	242
2	0.75	0.76	0.76	280
3	0.58	0.47	0.52	230
4	0.51	0.72	0.60	230
5	0.73	0.49	0.59	263
6	0.80	0.70	0.75	262
7	0.53	0.71	0.60	248
8	0.81	0.33	0.47	236
9	0.70	0.53	0.61	260
accuracy			0.61	2494
macro avg	0.64	0.61	0.60	2494
weighted avg	0.65	0.61	0.61	2494

Sejal Bhalla
2017100

2. With Noise: Noise samples were selected randomly and added after multiplying with a factor = 0.005 to each audio file, adding up to a total of 20000 files for training.

   Average Precision: **60%**
   Average Recall: **46%**
   Average F1 Score: **42%**
   Average Accuracy: **47%**

	precision	recall	f1-score	support
0	0.44	0.86	0.58	243
1	0.54	0.29	0.37	242
2	0.66	0.79	0.72	280
3	0.33	0.54	0.41	230
4	0.62	0.36	0.46	230
5	0.74	0.11	0.19	263
6	1.00	0.11	0.19	262
7	0.41	0.69	0.52	248
8	0.85	0.15	0.25	236
9	0.36	0.74	0.49	260
accuracy			0.47	2494
macro avg	0.60	0.46	0.42	2494
weighted avg	0.60	0.47	0.42	2494

Analysis

1. Augmenting the data with noise is supposed to increase the robustness and hence improve the performance. However, as opposed to this, adding noise either decreased or didn't change the Precision, Recall and Accuracy. This shows that the validation set was more similar to the training set and adding noise increased the variance by an amount larger than desired.

2. Spectrogram based features perform very poorly, with an accuracy less than chance. This shows that spectrograms are not able to capture the distinct features of an audio signal.

3. As expected, MFCC based features perform satisfactorily and much better than Spectrogram based features.

4. For Spectrograms, the extracted features are able to distinguish the sound of 'Six' (Precision = 95%) from all other sounds very effectively. This could indicate less variability in different sounds for 'Six'.

5. MFCC performs comparably well for all classes.