Assignment - 2
Saksham Vohra 2016085

Spectrograms

Procedure:

1. Zero pad all the audios to the same length(16000)
2. Slice the audio with window size = 256 and overlap of 84
3. Multiply the windows with a hanning window to force ends of the signal to zero and add distortion
4. Find FFT of the windows
5. Extract the real part and make them positive
6. Replace the 0 values with epsilon
7. Take Log of the signal to reduce sparsity

	precision	recall	f1-score	support
0	0.67	0.70	0.69	260
1	0.53	0.56	0.54	230
2	0.44	0.49	0.46	236
3	0.55	0.60	0.57	248
4	0.73	0.68	0.70	280
5	0.58	0.63	0.60	242
6	0.77	0.77	0.77	262
7	0.69	0.57	0.62	263
8	0.66	0.63	0.64	243
9	0.59	0.52	0.55	230
micro avg	0.62	0.62	0.62	2494
macro avg	0.62	0.61	0.62	2494
weighted avg	0.62	0.62	0.62	2494

0.6182838813151563

Data Augmentation

3 experiments were performed on the dataset

1. Add noise to 50% of audios
   Accuracy: 49.8%
2. Add noise to 30% of audios
   Accuracy: 54.2%
3. Add 30% noise to 50% audios
   Accuracy: 52.4%

It was observed that accuracy increased with decreasing the number of files corrupted

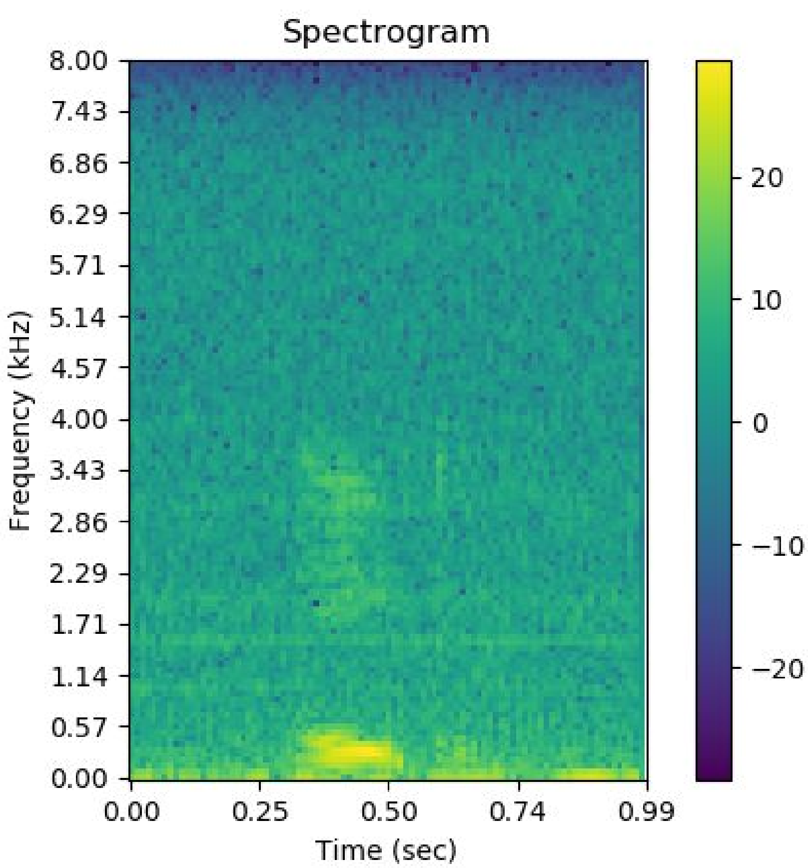
It was also observed that Linear SVM performed better than RBF kernel for the given task

During dataset analysis, it was also found that some sounds were too short to be heard by humans

	precision	recall	f1-score	support
0	0.54	0.53	0.54	260
1	0.46	0.49	0.47	230
2	0.35	0.47	0.40	236
3	0.44	0.47	0.45	248
4	0.63	0.66	0.65	280
5	0.52	0.57	0.54	242
6	0.74	0.57	0.64	262
7	0.57	0.51	0.54	263
8	0.59	0.51	0.55	243
9	0.46	0.43	0.44	230
micro avg	0.52	0.52	0.52	2494
macro avg	0.53	0.52	0.52	2494
weighted avg	0.54	0.52	0.53	2494

0.5232558139534884

Sample Plot



MFCC

Procedure:

1. Zero pad all the audios to the same length(16000)
2. Pre-emphasis is used on signal to increase the amplitude of higher frequency and reduce that of smaller frequencies
3. Slice the audio with window size = 256 and overlap of 84
4. Multiply the windows with a hanning window to force ends of the signal to zero and add distortion
5. Find FFT of the windows
6. Find the power spectrum to using formula in the paper
7. conver t the frequency from Hz to mel scale
8. Distribute signal evenly in mel scale accroess 80 filters
9. Convert distributed signal back to Hz
10. Create Filter banks and find DCT of the signal
11. Pass the resulting signal through a sin filter to increase the immunity to background noise

	precision	recall	f1-score	support
0	0.64	0.69	0.66	260
1	0.46	0.48	0.47	230
2	0.43	0.50	0.46	236
3	0.56	0.56	0.56	248
4	0.66	0.65	0.65	280
5	0.57	0.56	0.56	242
6	0.77	0.76	0.77	262
7	0.68	0.55	0.61	263
8	0.63	0.55	0.59	243
9	0.50	0.53	0.51	230
micro avg	0.59	0.59	0.59	2494
macro avg	0.59	0.58	0.58	2494
weighted avg	0.59	0.59	0.59	2494

Data Augmentation

3 experiments were performed on the dataset

4. Add noise to 50% of audios
   Accuracy: 44.2%
5. Add noise to 30% of audios
   Accuracy: 49.7%
6. Add 30% noise to 50% audios
   Accuracy: 46.8%

It was observed that accuracy increased with decreasing the number of files corrupted

It was also observed that RBF SVM performed better than linear kernel for the given task

During dataset analysis, it was also found that some sounds were too short to be heard by humans

	precision	recall	f1-score	support
0	0.47	0.57	0.51	260
1	0.36	0.44	0.40	230
2	0.31	0.38	0.34	236
3	0.40	0.35	0.37	248
4	0.55	0.54	0.54	280
5	0.53	0.55	0.54	242
6	0.62	0.61	0.61	262
7	0.53	0.44	0.48	263
8	0.49	0.42	0.45	243
9	0.44	0.35	0.39	230
micro avg	0.47	0.47	0.47	2494
macro avg	0.47	0.46	0.46	2494
weighted avg	0.47	0.47	0.47	2494

0.46792301523656776

MFCC Plots