**Niranjan Humagain**

**Machine Learning**

**Project-3**

**Q.1. FFT components for classification**

**i) Accuracy and Confusion Matrix:**

**1 2 3 4 5 6**

**1** *19 13 15 18 21 24*

**2** *9 19 14 22 14 22*

**3** *3 14 19 18 24 22*

**4** *4 15 14 26 20 21*

**5** *11 16 16 22 16 19*

**6** *6 14 14 19 18 29*

*Accuracy=19.67 %*

**ii) Bias:**

The accuracy was around 20%. The classification was bias in terms that the training and testing data were not shuffled. They were categorized according to pattern (taking modulo). This shows that arrangement of observations played some role in preventing to get optimal values for weights. Also training data were less enough to estimate parameters. Also, there are many ways of normalizing data’s among which I choose one way, which in this case seems not to be the best one.

**Q.2 MFCC components for classification**

1. **Accuracy and Confusion Matrix**

**1 2 3 4 5 6**

**1** *79 1 1 0 0 19*

**2** *12 9 2 15 0 62*

**3** *40 12 9 14 21 4*

**4** *14 7 0 76 1 2*

**5** *7 3 6 13 6 65*

**6** *1 1 0 1 38 59*

**Accuracy= 39.67%**

**ii) Bias**

We got accuracy of around 40%, which is much higher than what we got from fft. This shows mfcc is much efficient for classifying voices and music.

**Q.3 Best 20 Features Selection**

1. **Accuracy and Confusion Matrix:**

**1 2 3 4 5 6**

**1** *98 0 2 0 0 0*

**2** *98 0 1 0 1 0*

**3** *83 0 5 1 11 0*

**4** *82 0 10 3 5 0*

**5** *98 0 0 0 2 0*

**6** *83 0 10 2 5 0*

*Accuracy=18.00%*

**ii) Bias:**

I used the standard deviation for ranking best 20 features. Those with high standard deviation were given high ranks. But we can see that the result is very biased. Most of the music were categorized under Genre 1.No music’s were categorized under Genre 2 and Genre 6. This shows that standard deviation is not the good measure to rank our features. Also the normalization of data does not seem to be the best and I could have used other ways of normalization. Shuffling of data for training and testing could be done randomly and I guess it would produce better result.

**Q.4 Ways for improvements for classification**

i) Using features more than 1000 would increase the accuracy.

ii) Trying different ways of normalizing data and choosing the best one (based on accuracy level)

iii) Choosing the optimal values for constants like learning rate, regularization constants trying different values and choosing the best one (based on accuracy level)

iv) Shuffling data enough to categorize under test or trainee.