

Matthew Letter

Project 3

CS529 Machine Learning

April 9, 2015

### **Introduction:**

For this project, I was to Implement multinomial logistic regression and gradient descent for weight training based off Tom Mitchells book (1). The python programming language was used as it has nice built in matrix math functions with numpy. In total 3 questions were ask and will be referred to as part A, B, and C. In A I was asked to use the first 1000 points of a Fast Fourier Transformation (FFT) to train the network. In B I was asked to choose twenty features of the FFT for training based off some ranking metric. In C I was Asked to use the MFCC for training.

### **General Overview:**

For importing data I used the suggested function from the assignment paper and created a function that recurses from the running directory to all sub-directories looking for wav files and using the directory name, above the wave files, as the class name. Pickle was then used to write the transformed wave file data to disk, so as to avoid having to reprocess the data for every run. The pickle data is then reloaded and fed into the next stage.

I implemented a generalized function that takes a map of {class:[vectors]} and performs the multinomial logistic regression algorithm using gradient descent on the map. First it takes the map and splits it into training and testing for 10x cross validation then it proceeds to do 2000 epochs of training before moving onto the next cross validation testing and training set. The confusion matrix is fed into the next run and the accuracies for each run are stored in a list for later averaging.

### Part A:

The FFT was extracted from all the wav files and fed into the regression algorithm. 10-fold cross validation was then performed and the results are shown in table 1. The averaged 10-fold accuracy was 49.17% which was obtained after averaging the 10-fold cross validation with 2000 epochs per fold with eta set to 0.02 and lambda set to 0.0001. Notice country and metal had the worst prediction rates. In my opinion, since country is similar in sound to rock and pop it got mixed in with these buckets. Also, for the same reasons I think this is why metal got mixed in with the other classes as well.

### TRUE

|           | Classical | Country | Jazz | Metal | Pop | Rock |
|-----------|-----------|---------|------|-------|-----|------|
| Classical | 79        | 1       | 6    | 2     | 8   | 4    |
| Country   | 5         | 17      | 12   | 5     | 37  | 24   |
| Jazz      | 15        | 3       | 50   | 5     | 12  | 15   |
| Metal     | 5         | 2       | 4    | 28    | 43  | 18   |
| Pop       | 4         | 1       | 4    | 3     | 78  | 10   |
| Rock      | 2         | 5       | 6    | 16    | 28  | 43   |

**Table 1:** Column is the true value; Row is the predicted value. Table 1 is the 10-fold confusion matrix for Part A.

### Part B:

For choosing the top twenty features, I chose the twenty features that had the largest weights from part A's weight matrix by class weight vector. I had a significantly higher 10-fold cross validation accuracy than part A (with the same algorithm settings) of 67.5% with a better confusion matrix (see table 2). I believe this shows that the features that maximize the weights in the weight matrix contain the most information. Notice that classical, pop, and rock are getting confused, I believe this is because they have a similar sound (as mentioned in part A).

| TRUE      |           |         |      |       |     |      |
|-----------|-----------|---------|------|-------|-----|------|
|           | Classical | Country | Jazz | Metal | Pop | Rock |
| Classical | 54        | 1       | 13   | 1     | 9   | 22   |
| Country   | 8         | 72      | 9    | 4     | 4   | 3    |
| Jazz      | 9         | 11      | 66   | 1     | 11  | 2    |
| Metal     | 7         | 2       | 9    | 64    | 13  | 5    |
| Pop       | 3         | 1       | 3    | 2     | 89  | 2    |
| Rock      | 15        | 2       | 4    | 3     | 26  | 60   |

**Table 2:** Column is the true value; Row is the predicted value. Table 2 is the 10-fold confusion matrix for Part B.

For choosing the top twenty features, I also chose to rank by features that minimized the standard deviation. I had a significantly lower 10-fold cross validation accuracy over part A and part B using weights (with the same algorithm settings) of 34.83% with a terrible confusion matrix (see table 3). Even though this method was not as good as previous ones, the classification was still messing up on the same genres of music. Based on this result I would have to say that minimizing the standard deviation is not a good approach to feature selection.

|           | TRUE      |         |      |       |     |      |
|-----------|-----------|---------|------|-------|-----|------|
|           | Classical | Country | Jazz | Metal | Pop | Rock |
| Classical | 27        | 21      | 18   | 13    | 5   | 16   |
| Country   | 26        | 23      | 1    | 13    | 25  | 12   |
| Jazz      | 41        | 12      | 20   | 16    | 7   | 4    |
| Metal     | 1         | 10      | 10   | 48    | 26  | 5    |
| Pop       | 2         | 10      | 1    | 12    | 66  | 9    |
| Rock      | 12        | 15      | 6    | 22    | 20  | 25   |

Table 3: Column is the true value; Row is the predicted value. Table 3 is the 10-fold confusion matrix for Part B using standard deviation minimization.

### Part C:

In part C I took the GFCC data from each wav file using talkbox. This was run through the algorithm for training and testing. Table 4 shows the confusion matrix for a 10-fold cross validation set with 2000 epochs per fold with eta set to 0.02 and lambda set to 0.0001. Interestingly, even with a lower dimensionality matrix, the averaged 10-fold

accuracy was 61.67%. This accuracy was a significant improvement of ~10% over learning with the FFT vector. Notice again country music is being misclassified with multiple other classes. I believe this is because country must be very similar to most other classes except for metal in terms of tempo. Also, notice that rock and metal were getting mistaken for each other I believe this is for that same reason as country, having a similar tempo.

|           | TRUE      |         |      |       |     |      |
|-----------|-----------|---------|------|-------|-----|------|
|           | Classical | Country | Jazz | Metal | Pop | Rock |
| Classical | 78        | 2       | 17   | 1     | 1   | 1    |
| Country   | 14        | 24      | 29   | 7     | 12  | 14   |
| Jazz      | 18        | 4       | 60   | 8     | 3   | 7    |
| Metal     | 0         | 0       | 4    | 93    | 1   | 2    |
| Pop       | 2         | 3       | 5    | 1     | 84  | 5    |
| Rock      | 1         | 4       | 27   | 33    | 4   | 31   |

**Table 4:** Column is the true value; Row is the predicted value. Table 4 is the 10-fold confusion matrix for Part C.

### **Conclusion:**

In conclusion there are a few key features to take away from this assignment. 0-1 Normalization is very important for avoiding NAN's. The size of the input vector for training is not as important as the amount of information contained in each point of the vector (as shown by part B and C). This tells us that one of the most important parts of running multinomial logistical regression is determining if the data we feed into the algorithm is "good" in the sense that it gives us the information we want. Lastly, minimizing the standard deviation is not a good approach to feature selection.

I would like to point out that I believe this process could be improved by running part A, and then averaging the resulting weights for each fold's weight matrix and then determining the top twenty features for part B with the averaged weight matrix. This would remove any biased that may be passed on by using only one weight matrix from one fold.

### **References:**

[1] Mitchell, Tom M. Machine Learning. New York: McGraw-Hill, 1997. Print.