Multiclass classification of songs' genre using one-vs-all and multinomial logistic regression

Anonymous

***Abstract*—¡TO be Inserted after the rest of the report.¿ Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.**

***Index Terms*—IEEEtran, journal, LaTeX, paper, template.**

I. Introduction

Music, by no doubt, has become an imperative part of people's life. People prefer listening to their favorite genre of songs for more personalized experience. Because of the availability of tons of songs, listeners crave for more personalized experience while at the same time, they prefer non-static music for mood augmentation.

The project, in particular focuses on finding a pattern to identify the genre of songs. The identification pattern can help to build an online recommendation platform using exisiting databse of users so that listeners can have more customized experience while in search for new songs. During the entire project, we address the question that given a list of songs, how well can we analyze the audio to find a pattern that helps us identiy songs that belong to similar genre of music; which is to say that we construct a predictor h(x) for each genre (Y) using the features (x) of the songs and map it to a probability h(x). Simpler this tasks may seem for a certain type of songs (e.g. heavy metal vs classical), the classification problem can become difficult for other types (e.g.rock vs blues).

II. Data Analysis

The dataset used for the project is a custom subset of the Million Song Dataset [1], and the labels were obtained from AllMusic.com [give ref.]. For simplicity, each song has been assigned only one label that corresponds to the most representative genre. The 10 labels are :

1) Pop Rock
2) Electronic
3) Rap
4) Jazz
5) Latin
6) RnB
7) International
8) Country
9) Reggae
10) Blues

The layout of a single example, representing one song, is as follows:

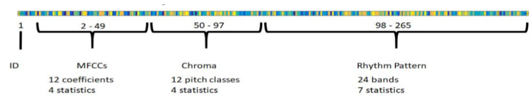

Fig. 1. Layout of Example Vector

The feature vector for each song thus consists of 12 coefficients representing MFFCs (using 4 statistics each), 12 pitch classes representing Chroma (using 4 statistics each), and 24 bands representing Rythm Patterns (using 7 statistics each). The details on the structure of each component of the example vector is as below:

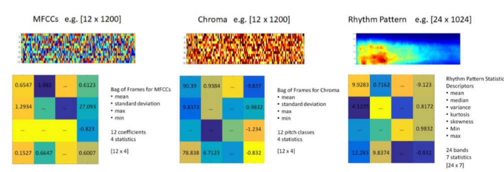

Fig. 2. Structural Representation of Example Vector

The covariance matrix of the 264 features provided in the dataset is as follows:

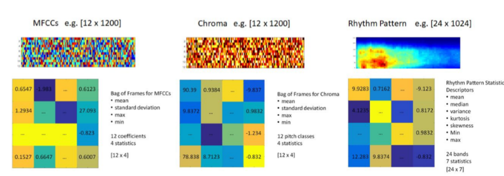

Fig. 3. Structural Representation of Example Vector

It is evident that the example features have high dimensionality, so an unsupervised learning approach was chosen to select create a lower dimensional dataset out of the existing dataset, to improve performance and eliminate potential redundancy in the dataset. The approach adopted for this was Principal Component Analysis: it maps the data to a lower dimensional space such that the variance of this lower dimensional representation is maximized.

The principal components obtained thus (i.e. the eigen vectors corresponding to the largest eigenvalues), may now be employed to rebuild a large part of the variance of the original data. Note that the principal components are orthogonal to each other, with the first principal component explaining the largest part of the variance of the original data, the second less than that, and so on.

For intuition, each principal component may be considered **one** of the new features of each example vector, that is now uncorrelated to the rest of the features (i.e. other principal components). While reducing the dimensions does lead to some data loss, it is restricted to the eigen vectors with lowest variance (i.e. those that don't have significant impact on the dynamics of the original dataset).

The covariance of the reduced dimension dataset illustrates this:

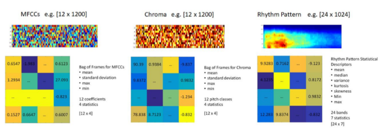

Fig. 4. Structural Representation of Example Vector

III. Methods and Experiments

Multiclass classification problem was explored using two different methods:

- Multinomial logistic regression
- One-vs-All (OvA or OVR)

Multinomial Logistic Regression is the linear regression analysis technique used when the dependent variable is nominal with more than two levels. Thus it is an extension of logistic regression, which analyzes dichotomous (binary) dependents. Standard linear regression requires the dependent variable to be of continuous-level (interval or ratio) scale. Logistic regression jumps the gap by assuming that the dependent variable is a stochastic event. And the dependent variable describes the outcome of this stochastic event with a density function (a function of cumulated probabilities ranging from 0 to 1).

The idea is to construct a linear predictor function that constructs a score from a set of weights that are linearly combined with the explanatory variables (features) of a given observation using a dot product:

$$score(X_i, k) = \beta_k.X_i, \tag{1}$$

where X_i is the vector of explanatory variables describing observation i, β_k is a vector of weights (or regression coefficients) corresponding to outcome k, and $score(X_i, k)$ is the score associated with assigning observation i to category k.

In the multinomial logit model, for K possible outcomes, K-1 independent binary logistic regression models are run, in which one outcome is chosen as a "pivot" and then the other K-1 outcomes are separately regressed against the pivot outcome.

Therefore, for each posssible outcome, we can write down that:

$$Equationstobeinsertedfromwiki \tag{2}$$

$$score(X_i, k) = \beta_k.X_i, \tag{3}$$

$$score(X_i, k) = \beta_k.X_i, \tag{4}$$

we can write down the expressions to for k classes as:

$$P(Y_i = 1) = \tag{5}$$

The core algorithm of the project relies on Multinomial Logistic Regression is the linear regression analysis to conduct when the dependent variable is nominal with more than two levels. Thus it is an extension of logistic regression, which analyzes dichotomous (binary) dependents. Since the SPSS output of the analysis is somewhat different to the logistic regressions output, multinomial regression is sometimes used instead.

The second approach explored for this project was One-vs-All (also called One-vs-Rest) Classification. We train one classifier for every class, taking the samples for that particular class as positive, while the remaining samples are treated as negative.

The base classifer must yield a real-valued confidence score for the decision, as opposed to merely a class label, since discrete class labels could become ambiguous if results from multiple classifiers yield different labels for the same sample.

An algorithm describing the procedure is as follows:

- *Inputs to the algorithm:*
  - dataset of samples X and labels y, where $y_i \in \{1, 2, ..., N\}$ is the label for the particular sample X_i
  - L, a binary classification training algorithm
- *Outputs of the algorithm:*
  - a set of classifiers c_n for $n \in \{1, 2, ..., N\}$
- *Mechanism:*
  - For every $n \in \{1, 2, ..., N\}$
    * Build a separate label vector z, with $z_i = 1$ if $y_i = n$ and $z_i = 0$ for all other y_i
    * Apply training algorithm L on samples X and label vector z, to get c_n

Concretely deciding which class an unseen sample x belongs to requires the application of all the obtained classifiers c_n to it, and then choosing the label n whose classifier yields the largest confidence score:

$$\hat{y} = \underset{n \in \{1,2,...,N\}}{argmax} \; c_n(x)$$

Two potential issues with the OVR are approach are:

- The confidence values can be different in terms of scale between each binary classifier
- Despite the training set having a balanced class distribution, the binary classifiers will see it as unbalanced since

the number of negative labels will (most likely) be much greater than the number of positive labels.

In our implementation of the OVR algorithm, we used a 2nd degree RBF Support Vector Classifier as our binary learner, but the results were practically unchanged from the first approach i.e. Multinomial Logistic Regression.

IV. Results

FOR ACCURACY : Entry No. 1 Using multinomial : Test Result : Train Result: Result on Kaggle:

Entry No.2 Using mutinomial: Test Result: Train Result: Result on Kaggle:

Test No. 3 Using OVR : Train Accuracy : 0.7753 Test Accuracy : 0.646294881589 Result on Kaggle : 0.62882 (Improvement)

FOR LOG-LOSS Entry No. 1: 2.60627 (using multinomial)

V. Conclusion

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

VI. Appendices

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Some Appendix The contents...

A Proof of the First Zonklar Equation

Some text for the appendix.

Acknowledgment

The authors would like to thank...

References