Music Genre Classification with Machine Learning

Li Guo, Zhiwei Gu, Tianchi Liu

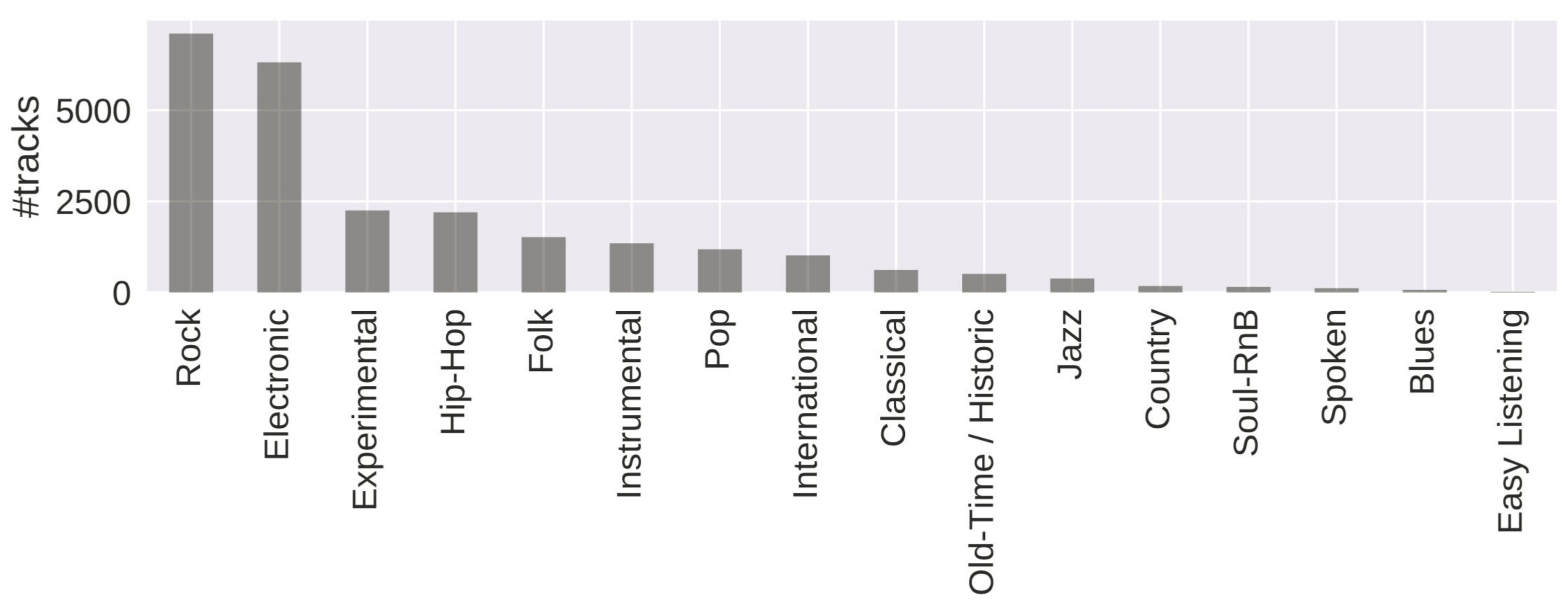
{liguo94, zhiweig, kitliu5}@stanford.edu

Objective

- Build effective music genre classification models using a variety of machine learning techniques
- Accurately classify genre of new music tracks with associated features

Dataset

Our Free Music Archive (FMA) dataset includes 106,574 tracks of music, splitted into 16 different genres, with 518 associated features extracted with LibROSA and Echonest



feature	chroma_cens										...
statistics	kurtosis										...
number	01	02	03	04	05	06	07	08	09	10	...
track_id											
2	7.180653	5.230309	0.249321	1.347620	1.482478	0.531371	1.481593	2.691455	0.866868	1.341231	...
3	1.888963	0.760539	0.345297	2.295201	1.654031	0.067592	1.366848	1.054094	0.108103	0.619185	...
5	0.527563	-0.077654	-0.279610	0.685883	1.937570	0.880839	-0.923192	-0.927232	0.666617	1.038546	...
10	3.702245	-0.291193	2.196742	-0.234449	1.367364	0.998411	1.770694	1.604566	0.521217	1.982386	...
20	-0.193837	-0.198527	0.201546	0.258556	0.775204	0.084794	-0.289294	-0.816410	0.043851	-0.804761	...

Preprocessing

- Conversion of audio raw data to 518 audio features using Python package LibROSA
- Splitting of all 25000 tracks into training set, validation set and test set, each with size of 19922, 2505, 2573
- Random shuffling of all training data

Models

1. Softmax

$$p(y=i|x;\theta)=\phi_i=\frac{\exp(\theta_i^T x)}{\sum_{j=1}^{k}\exp(\theta_j^T x)}$$

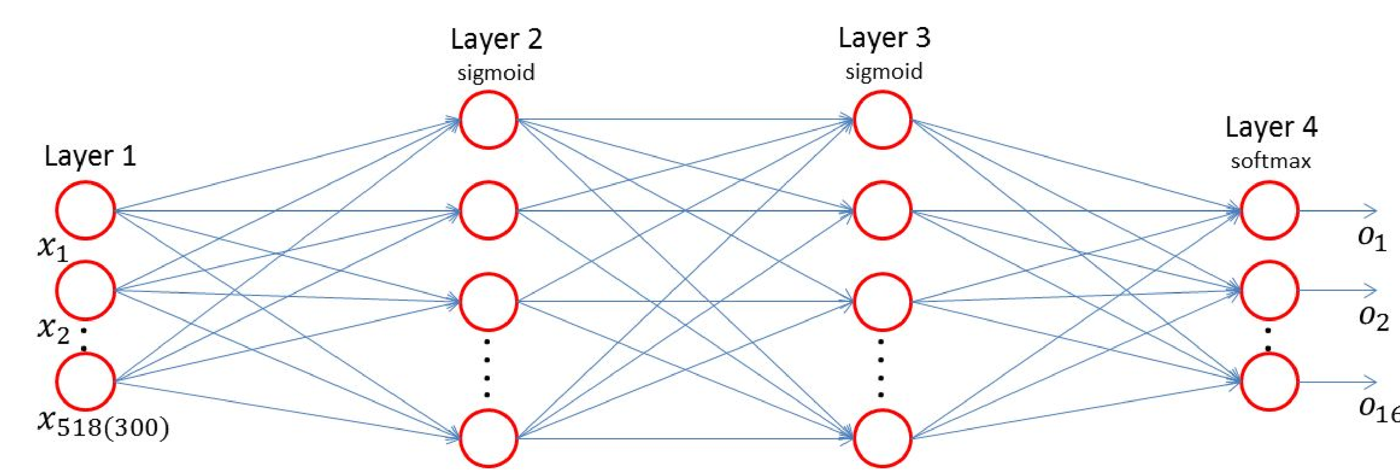
2. SVM-RBF

$$\min_{\gamma,w,b}\ \frac{1}{2}||w||^2$$

$$s.t.\ \ y^{(i)}(w^T+x^{(i)}+b)\geq\hat{\gamma},\quad i=1,...,m$$

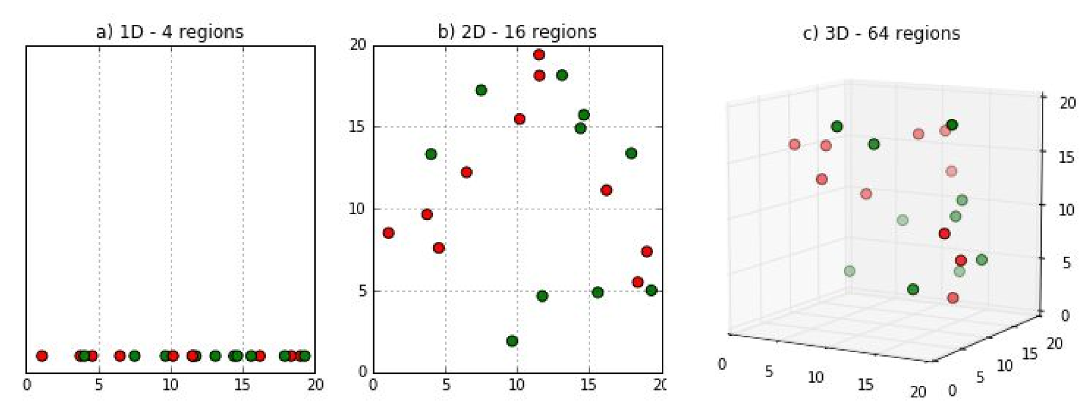
$$L_{hinge}(z,y)=\max(0,1-yz)\qquad K(x,z)=\exp\left(-\frac{1}{2}||x-z||^2\right)$$

3. Neural Network



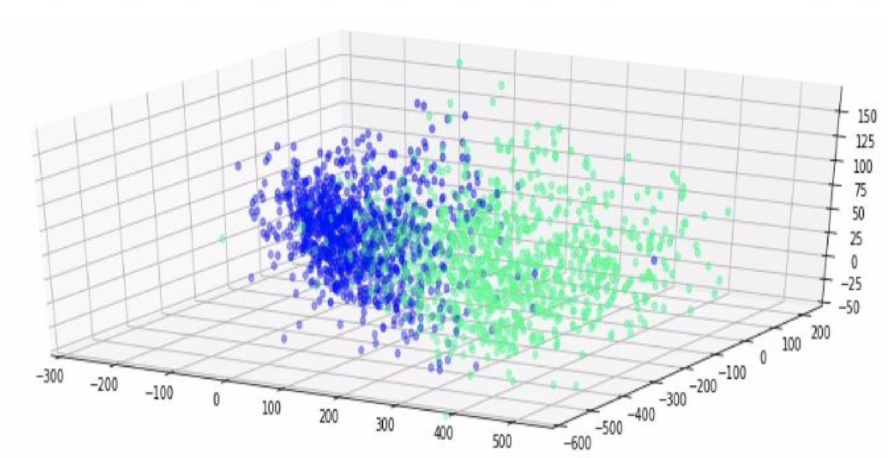
Neural Networks with sigmoid (layer 2 and 3) and softmax (layer 4) as activation functions. Features selection applied for overfitting reduction

4. KNN-PCA/Model Selection



<u>Curse of dimensionality</u>

sparse neighbors in very high dimension due to exponential volume expansion

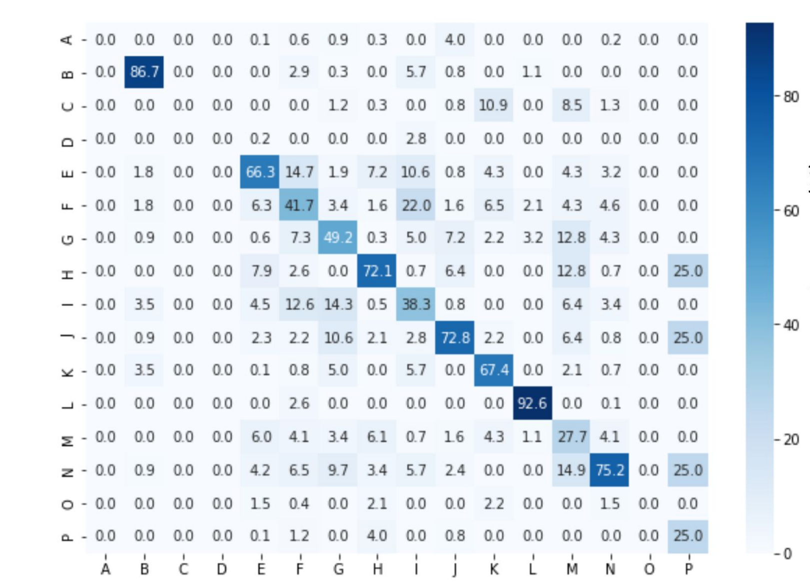


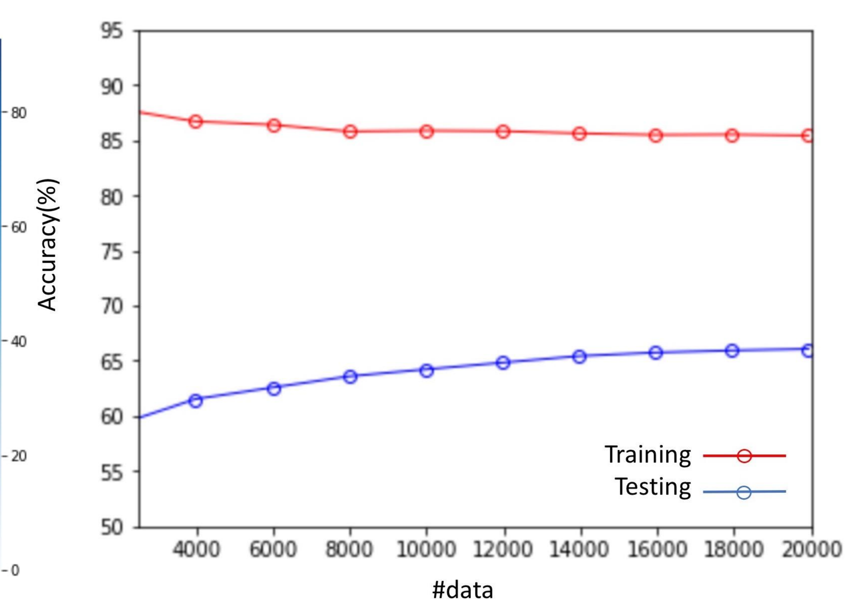
<u>Dimensionality reduction</u>

1. *Best subset feature selection*
   Closer neighbors and reduced model complexity
2. *Principal Components*
   Closer neighbors and better cluster segregation

Results

Models	Train accuracy	Test accuracy
Softmax	52.87%	51.03%
Logistic Regression	67.45%	62.61%
Neural Network	74.93%	63.19%
SVM Linear	67.38%	61.48%
SVM RBF	85.4%	66.07%
KNN-Model Selection	99.98%	57.87%





- SVM with RBF kernel gains test accuracy of 66.07%, increasing 30% from baseline, Softmax model;
- Improvement in test accuracy by L1 regularization and best subset feature selection

Future work

- **Model ensembling**: combining classifiers by voting or averaging to improve performance
- **Feature refining**: add other musically relevant features for better classification results
- **Real application**: new music tracks can turn into features the same way as we mentioned, and applied our machine learning models to predict its genre.

References

[1] Benzi, Kirell, et al. "FMA: A Dataset For Music Analysis." *arXiv preprint arXiv:1612.01840* (2016)

[2] Lee, Chang-Hsing, et al. "Automatic music genre classification using modulation spectral contrast feature." Multimedia and Expo, 2007 IEEE International Conference on. IEEE, 2007