Music Genre Classification

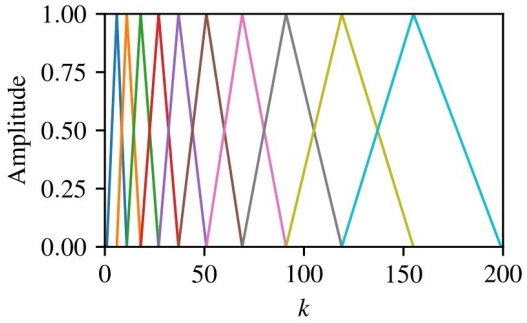
Audio Spectrum Analysis

Problems to Address

- High magnitude of content
- Artist labelling
  - Audience perspective
  - Verification
  - Bad actors
- Recommendations

The Data

- 1000 thirty second samples recorded and labelled in 2000-2001
- Further broken down to 3 second samples
- Full audio spectrum analysis
  - Spectral centroid, rolloff, etc.
- Divided spectral analysis
  - Mel Frequency Cepstral Coefficients



	filename	rolloff_mean	mfcc1_mean
0	blues.00000.wav	3805.839606	-113.570648

	filename	rolloff_mean	mfcc1_mean
0	blues.00000.0.wav	3714.560359	-118.627914
1	blues.00000.1.wav	3869.682242	-125.590706
2	blues.00000.2.wav	3997.639160	-132.441940
3	blues.00000.3.wav	3568.300218	-118.231087
4	blues.00000.4.wav	3469.992864	-105.968376
5	blues.00000.5.wav	4371.985614	-100.752792
6	blues.00000.6.wav	4325.026668	-101.773033
7	blues.00000.7.wav	3625.280386	-109.165077
8	blues.00000.8.wav	3586.934721	-113.373199
9	blues.00000.9.wav	3505.522649	-125.532906

Feature Selection

- Nearly 60
- Balanced deceptive multicollinearity
- Feature importances
- Model coefficients
- Manual Tuning
- 36 in final
  - Mainly Mel Frequency Coefficients
  - No PCA

VIF results

const	2703.515224
chroma_stft_mean	4.118726
chroma_stft_var	2.526067
rms_mean	51.227912
rms_var	3.840067
spectral_centroid_mean	185.494010
spectral_centroid_var	20.730610
spectral_bandwidth_mean	77.964079
spectral_bandwidth_var	6.843537
rolloff_mean	118.856952
rolloff_var	15.139926
zero_crossing_rate_mean	26.767622
zero_crossing_rate_var	8.054137
harmony_mean	1.487883
harmony_var	15.550854
perceptr_mean	1.595504
perceptr_var	9.233448
tempo	1.014670
mfcc1_mean	19.701721
mfcc1_var	2.138339
mfcc2_mean	20.273645
mfcc2_var	4.009512
mfcc3_mean	5.610340
mfcc3_var	2.204361
mfcc4_mean	3.131581
mfcc4_var	2.203961
mfcc5_mean	2.955767
mfcc5_var	2.150134

Model Selection

- KNN / Gradient Boost / Random Forest
  - Unavoidable Overfitting
- SVM
  - Linear only viable option
- Logistic Regression
  - Model simplicity
  - Explanatory power

Model Tuning

- GridSearchCV
  - Regularization
  - C values
  - Solvers
- ElasticNet
  - C = 0.1
  - Solver = ‘saga’
  - L1 ratio = 0.5

Results

Testing Data

	precision	recall	f1-score	support
blues	0.74	0.61	0.67	200
classical	0.83	0.97	0.89	200
country	0.61	0.56	0.58	200
disco	0.60	0.64	0.62	200
hiphop	0.58	0.47	0.52	200
jazz	0.69	0.68	0.69	200
metal	0.71	0.85	0.77	200
pop	0.67	0.82	0.74	200
reggae	0.63	0.63	0.63	200
rock	0.34	0.28	0.31	200
accuracy			0.65	2000
macro avg	0.64	0.65	0.64	2000
weighted avg	0.64	0.65	0.64	2000

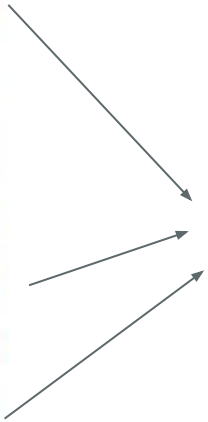
Training Data

	precision	recall	f1-score	support
blues	0.62	0.69	0.65	800
classical	0.91	0.93	0.92	798
country	0.62	0.58	0.60	797
disco	0.62	0.61	0.61	799
hiphop	0.70	0.68	0.69	798
jazz	0.78	0.79	0.78	800
metal	0.81	0.87	0.84	800
pop	0.74	0.80	0.77	800
reggae	0.69	0.63	0.66	800
rock	0.54	0.47	0.50	798
accuracy			0.71	7990
macro avg	0.70	0.71	0.70	7990
weighted avg	0.70	0.71	0.70	7990

Ten fold cross validation: 0.59

Prediction Voting

	filename	predictions
40	blues.00004.0.wav	classical
41	blues.00004.1.wav	reggae
42	blues.00004.2.wav	blues
43	blues.00004.3.wav	country
44	blues.00004.4.wav	classical
45	blues.00004.5.wav	disco
46	blues.00004.6.wav	classical
47	blues.00004.7.wav	blues
48	blues.00004.8.wav	blues
49	blues.00004.9.wav	country



	filename	label	vote_pred	model_pred	avg_vote
4	blues.00004.wav	blues	classical	country	blues

Probability Aggregation

	filename	blues	classical	country	disco	hiphop	jazz	metal	pop	reggae	rock
40	blues.00004.0.wav	0.06	0.51	0.16	0.04	0.00	0.07	0.0	0.0	0.01	0.15
41	blues.00004.1.wav	0.27	0.00	0.18	0.11	0.02	0.00	0.0	0.0	0.30	0.12
42	blues.00004.2.wav	0.55	0.00	0.14	0.09	0.04	0.00	0.0	0.0	0.06	0.10
43	blues.00004.3.wav	0.14	0.01	0.45	0.11	0.01	0.01	0.0	0.0	0.13	0.14
44	blues.00004.4.wav	0.21	0.32	0.08	0.22	0.01	0.10	0.0	0.0	0.02	0.04
45	blues.00004.5.wav	0.17	0.01	0.18	0.34	0.05	0.01	0.0	0.0	0.08	0.17
46	blues.00004.6.wav	0.05	0.59	0.08	0.09	0.00	0.12	0.0	0.0	0.01	0.06
47	blues.00004.7.wav	0.43	0.03	0.12	0.17	0.06	0.01	0.0	0.0	0.11	0.06
48	blues.00004.8.wav	0.26	0.06	0.23	0.19	0.01	0.05	0.0	0.0	0.06	0.15
49	blues.00004.9.wav	0.21	0.10	0.30	0.05	0.01	0.02	0.0	0.0	0.05	0.28

	4
filename	blues.00004.wav
avg_vote	blues
blues	0.24
classical	0.16
country	0.19
disco	0.14
hiphop	0.02
jazz	0.04
metal	0
pop	0
reggae	0.08
rock	0.13

Aggregation Results

Testing Data
0.64

Voting

	precision	recall	f1-score	
accuracy			0.70	200
macro avg	0.69	0.71	0.69	200
weighted avg	0.69	0.70	0.69	200

Probability Aggregation

	precision	recall	f1-score	
accuracy			0.71	200
macro avg	0.70	0.71	0.70	200
weighted avg	0.70	0.71	0.70	200

Model Prediction

	precision	recall	f1-score	
accuracy			0.68	200
macro avg	0.67	0.68	0.67	200
weighted avg	0.67	0.68	0.67	200

Training Data
0.70

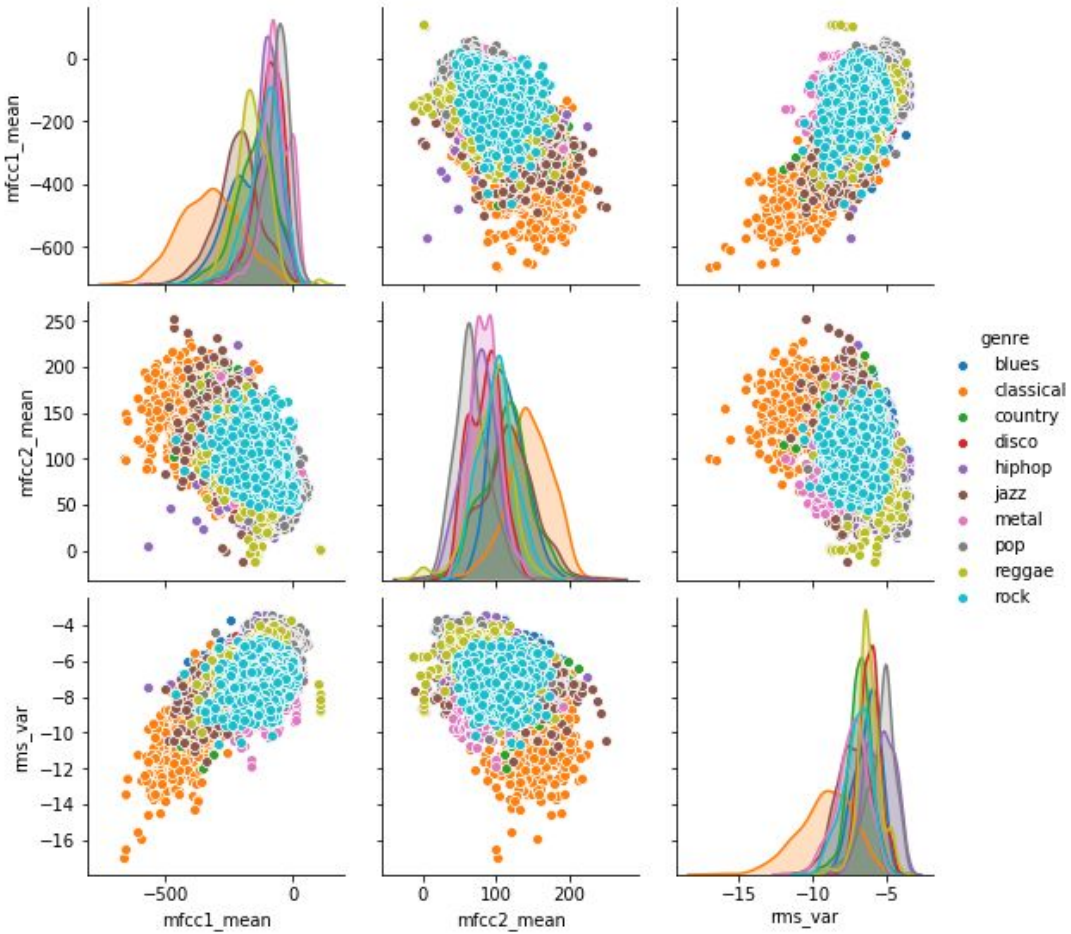
	precision	recall	f1-score	
accuracy			0.80	800
macro avg	0.80	0.80	0.80	800
weighted avg	0.80	0.80	0.80	800

	precision	recall	f1-score	
accuracy			0.81	800
macro avg	0.81	0.81	0.81	800
weighted avg	0.81	0.81	0.81	800

	precision	recall	f1-score	
accuracy			0.74	800
macro avg	0.75	0.74	0.74	800
weighted avg	0.75	0.74	0.74	800

Feature Analysis

	feat	coef	abs_coef
9	mfcc1_mean	-1.441043	1.441043
10	mfcc2_mean	1.137316	1.137316
2	rms_var	1.039597	1.039597
6	harmony_var	1.028325	1.028325
14	mfcc4_mean	0.913630	0.913630
3	zero_crossing_rate_mean	0.832232	0.832232
17	mfcc6_mean	0.827381	0.827381
11	mfcc2_var	-0.635520	0.635520
12	mfcc3_mean	0.508446	0.508446
19	mfcc7_mean	-0.503555	0.503555
16	mfcc5_var	0.475959	0.475959
33	mfcc18_mean	-0.436142	0.436142
20	mfcc8_mean	0.417501	0.417501



Confusion Matrix

Predicted

Actual

	blues	classical	country	disco	hiphop	jazz	metal	pop	reggae	rock
blues	122	0	9	3	0	24	10	0	20	12
classical	0	194	1	0	0	4	0	0	0	1
country	17	0	111	6	9	7	6	3	5	36
disco	0	1	6	128	19	1	3	21	2	19
hiphop	2	0	8	13	93	0	10	36	33	5
jazz	1	35	12	1	0	136	0	1	3	11
metal	0	0	0	11	8	0	171	0	3	7
pop	0	1	7	4	12	4	0	164	3	5
reggae	6	0	17	12	14	0	8	8	126	9
rock	17	3	12	37	5	20	34	13	4	55

Song Recommendations

filename	blues.00004.wav
avg_vote	blues
blues	0.24
classical	0.16
country	0.19
disco	0.14
hiphop	0.02
jazz	0.04
metal	0
pop	0
reggae	0.08
rock	0.13

- Individual user taste profiles
- Goal priorities
  - Better listening experience
  - Accurate predictions
- A/B testing for genre tolerance
- Other classifiers

Conclusions

- Difficult to avoid data leakage
- Genre Classification Probabilities
- Moving forward
  - How does it handle more classes?
  - Does it improve with higher density analysis?
  - Are genres our best classes?

Questions?

Thank You