Music Genre Classification using Naive Bayes and Logistic Regression

Does subsampling the best features affect model performance?

Matt Gusdorff
Keith Mburu

How do we identify music?

30 seconds

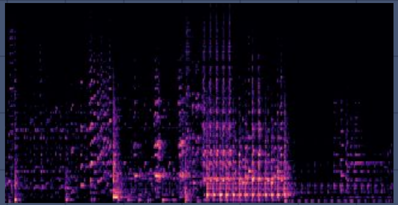
3 seconds

Motivation

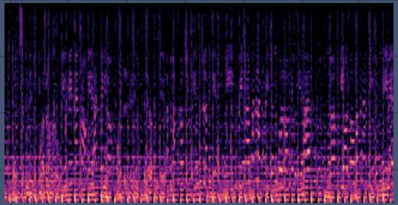
To investigate the objective factors (features) that determine the genre of a piece of music, to identify which of these are most predictive of it, and to establish whether including those that are less predictive can introduce noise that makes classification more difficult

Can only using a sample of the best features of a three second song clip make it easier to predict what genre the song is?

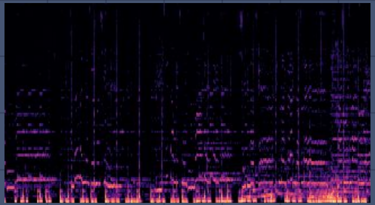
Dataset



Jazz



Rock



Pop

filename	tempo	harmony_	spectral_centroid_mean	rolloff_var	mfcc1_mean
metal.00046.3.wav	117.4538	3.70E-06	2901.964416	156037.83	6.813441753
country.00038.9.wav	151.9991	-0.00016	2928.853378	1298035.1	-132.388504
country.00023.1.wav	80.74951	2.79E-05	1504.244447	4482771.6	-379.795441
rock.00001.9.wav	78.30256	1.47E-05	1759.781989	1900912.3	-153.797836
pop.00052.6.wav	103.3594	-4.52E-05	3707.800394	2027995.6	-44.3918724
reggae.00066.3.wav	80.74951	-6.72E-05	1676.874206	1957733.8	-145.4767
classical.00019.4.wav	123.0469	-3.33E-07	1010.957411	418535.04	-546.663818
rock.00053.4.wav	117.4538	0.0009	3236.201432	886051.03	7.388467312
classical.00019.8.wav	184.5703	-6.79E-06	958.7397532	86032.069	-402.166351

https://www.kaggle.com/andradaolteanu/gtzan-dataset-music-genre-classification

Dataset

57 continuous features, including:

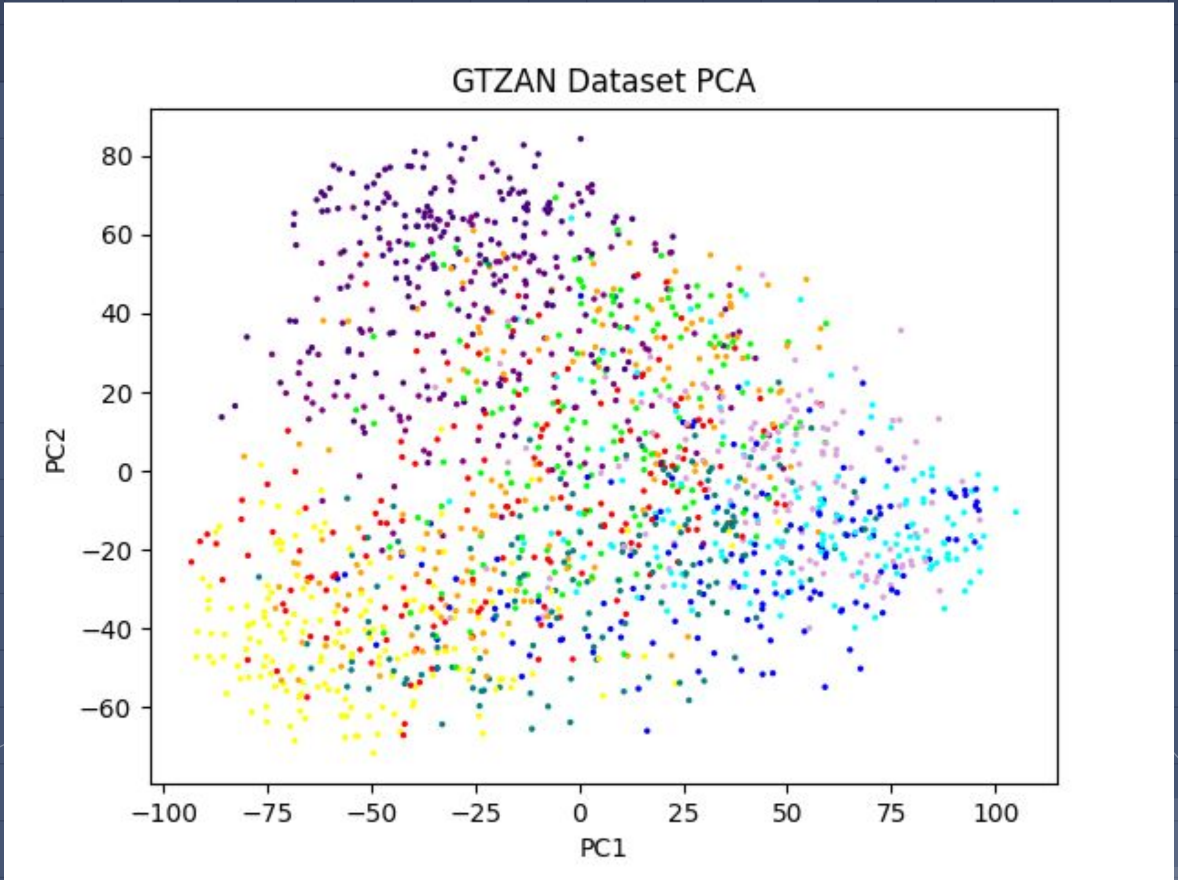
- Chroma_stft mean, variance
- Rms mean, variance
- Spectral_bandwidth mean, variance
- Zero_crossing_rate mean, variance
- Perceptr mean, variance

Translate to ~380,000 discrete features

Labels:

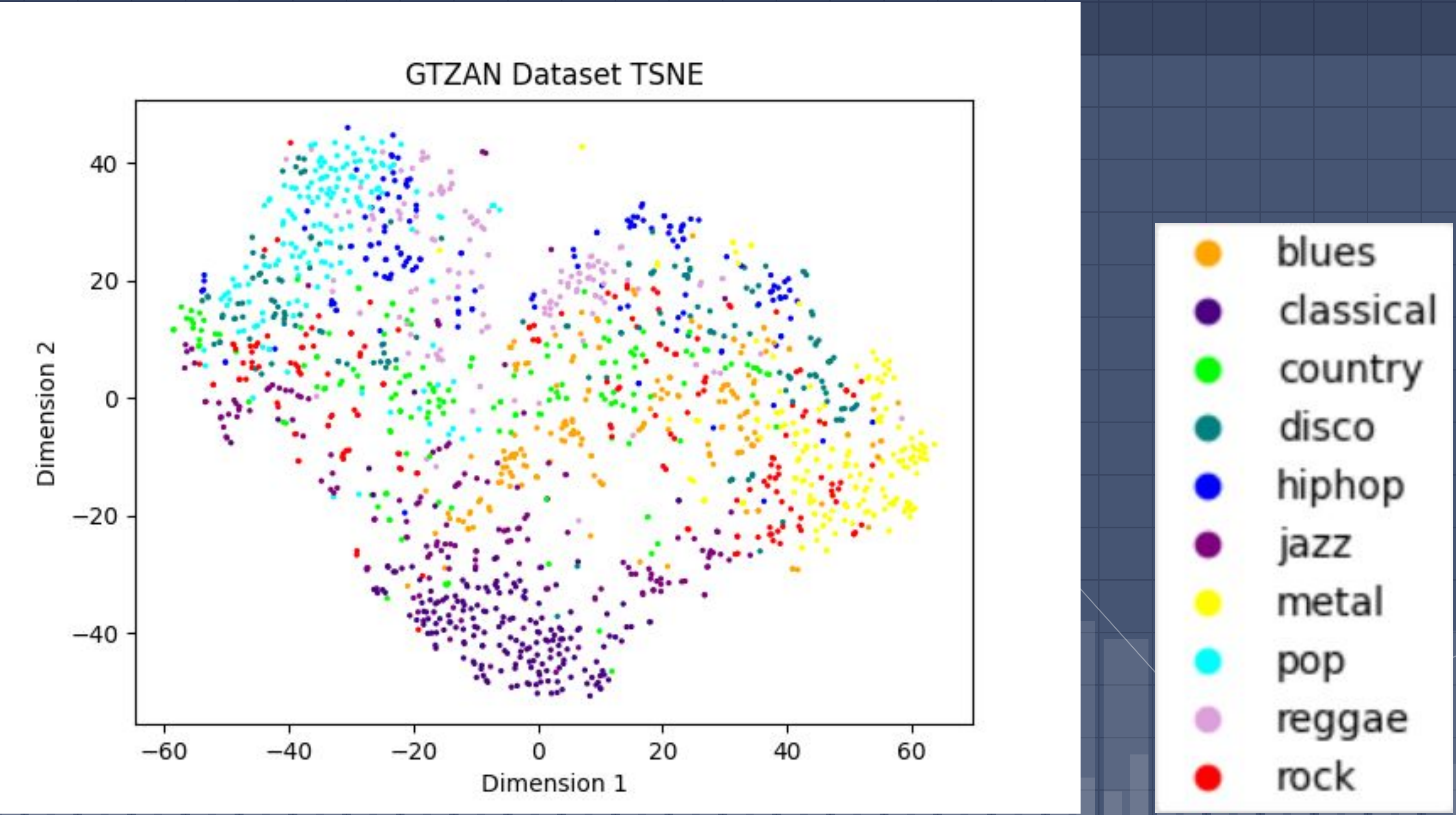
- Rock
- Blues
- Metal
- Pop
- Country
- Classical
- Jazz
- Reggae
- Hiphop
- Disco

Dimensionality Reduction : PCA



Explained variance ratios:
PC1 - 0.14886974
PC2 - 0.09551203

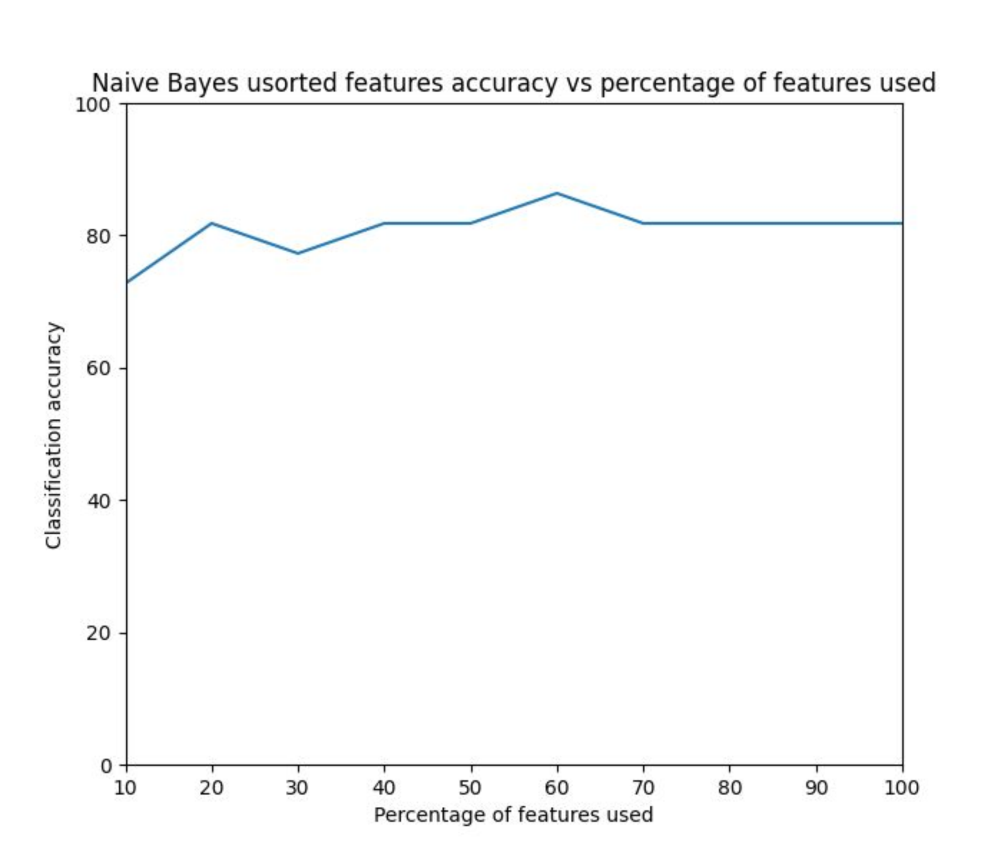
Dimensionality Reduction : t-SNE



Naive Bayes

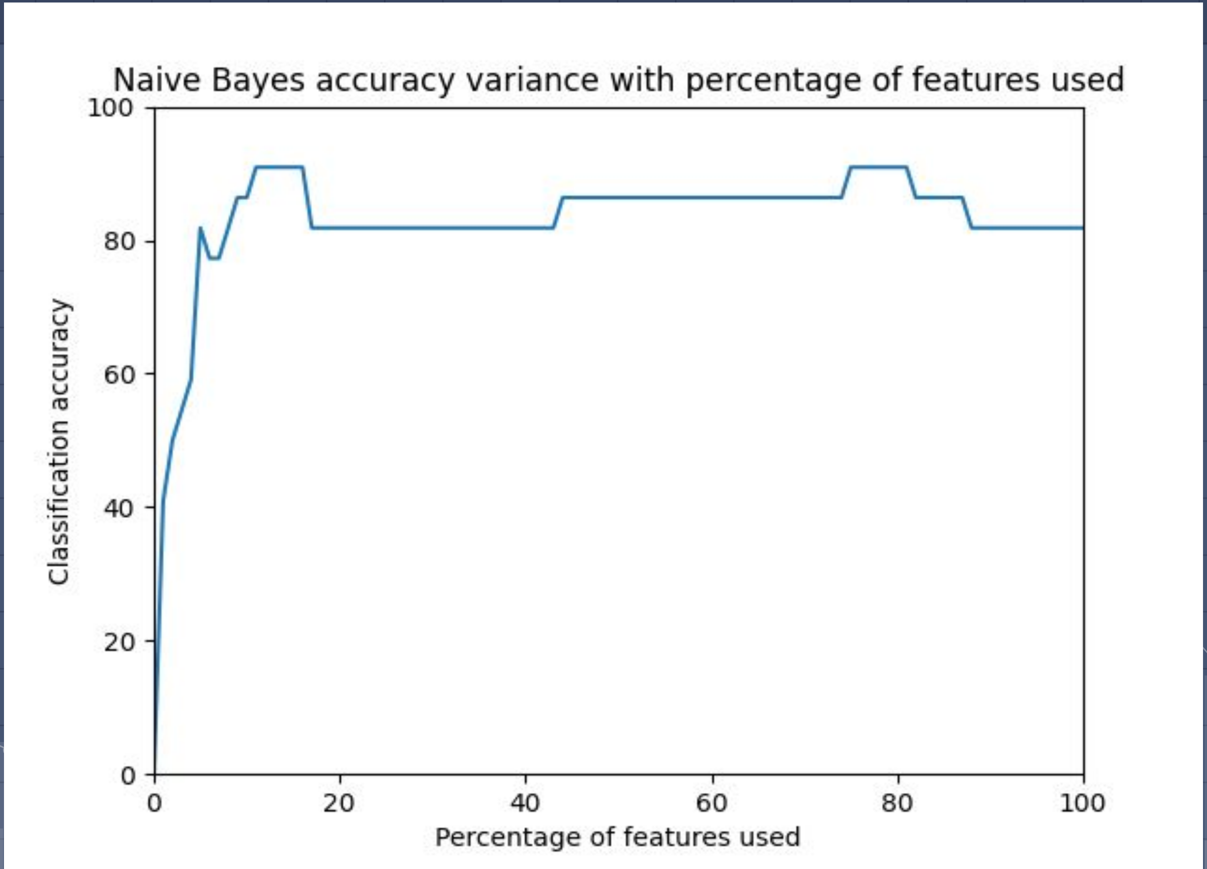
- Using discrete features converted from continuous features.
- Evaluation with different proportions of features included.
- Best features determined using entropy

Naive Bayes unsorted features accuracy



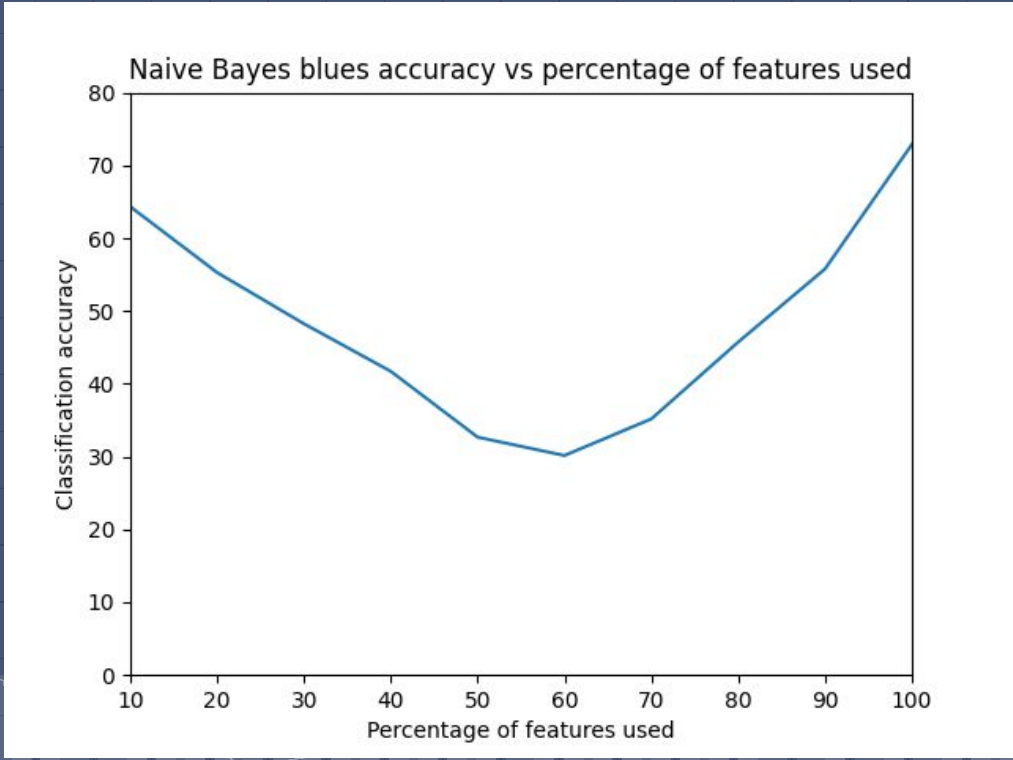
- Generally, more features = better

Naive Bayes best features accuracy



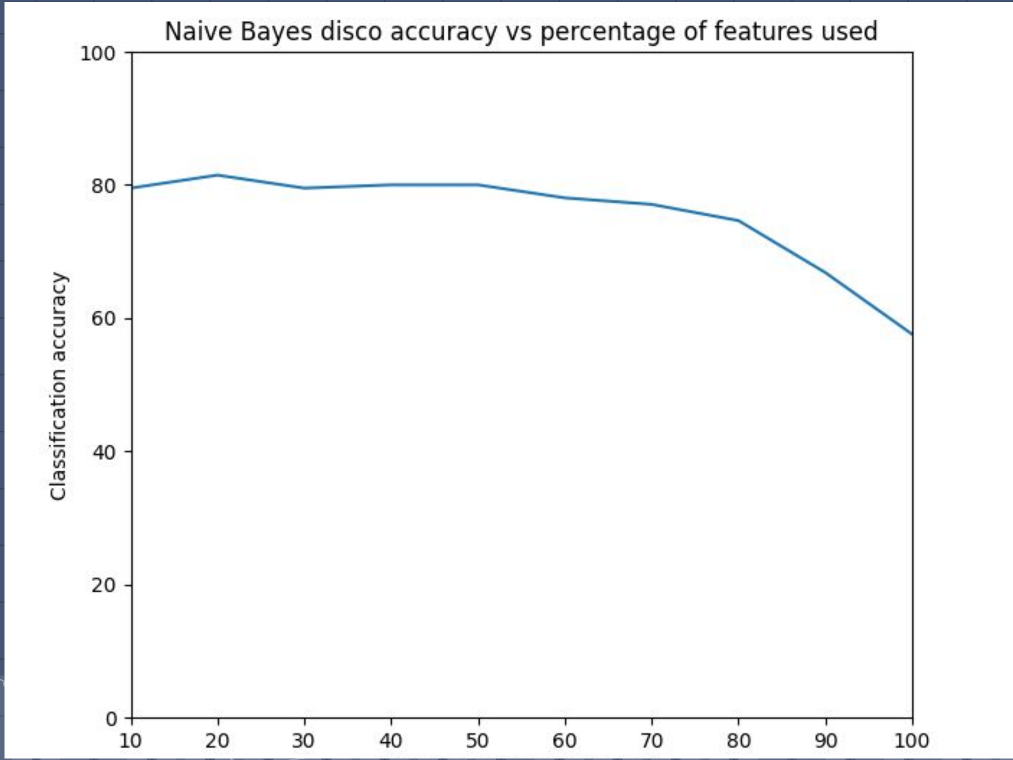
- Best features prioritized for feature selection
- Diminishing returns

Naive Bayes Blues accuracy



	Blues		Country			Jazz				Rock
blues	128	1	56	3	1	9	0	0	1	0
blues	110	1	46	2	0	32	0	0	0	8
blues	96	1	44	1	0	43	0	0	2	12
blues	83	1	53	1	0	44	0	0	2	15
blues	65	1	59	1	0	57	0	0	3	13
blues	60	1	58	1	0	64	0	0	3	12
blues	70	1	57	1	0	54	0	0	2	14
blues	91	1	52	0	1	36	0	0	2	16
blues	111	3	32	0	0	33	0	0	1	19
blues	145	3	9	0	0	33	0	0	1	8

Naive Bayes Disco accuracy



				Disco	Hiphop			Pop		Rock
disco	1	0	3	163	10	0	0	22	3	3
disco	0	0	1	167	7	1	0	20	0	9
disco	0	0	1	163	6	1	0	25	0	9
disco	0	0	1	164	8	1	0	22	0	9
disco	0	0	1	164	7	1	0	21	1	10
disco	0	0	1	160	8	1	0	21	1	13
disco	1	0	1	158	8	1	0	20	1	15
disco	1	0	2	153	9	0	0	20	1	19
disco	6	0	3	137	11	0	0	23	0	25
disco	13	0	2	118	15	0	0	31	0	26

Logistic Regression

- Using the original continuous features

- Stochastic Gradient Descent to find weights

$$h_{\boldsymbol{w}}(\boldsymbol{x}) = p(y = 1|\boldsymbol{x}) = \frac{1}{1 + e^{-\boldsymbol{w}\cdot\boldsymbol{x}}}$$

$$J(\boldsymbol{w}) = -\sum_{i=1}^{n} y_i \log h_{\boldsymbol{w}}(\boldsymbol{x_i}) + (1 - y_i)\log(1 - h_{\boldsymbol{w}}(\boldsymbol{x_i}))$$

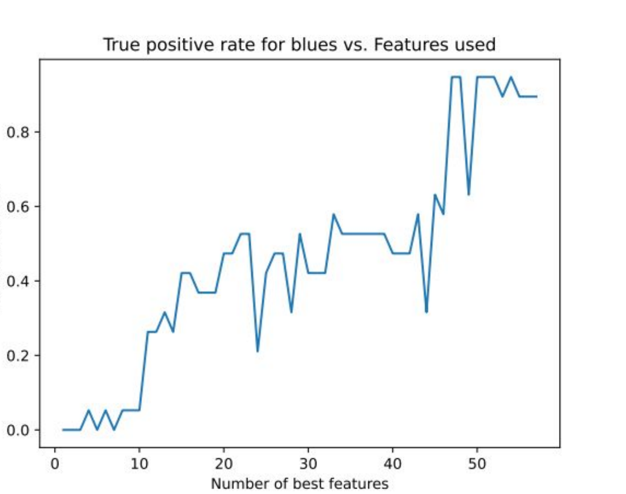
$$\nabla J_{\boldsymbol{x}_i}(\boldsymbol{w}) = (h_{\boldsymbol{w}}(\boldsymbol{x_i}) - y_i)\boldsymbol{x_i}$$

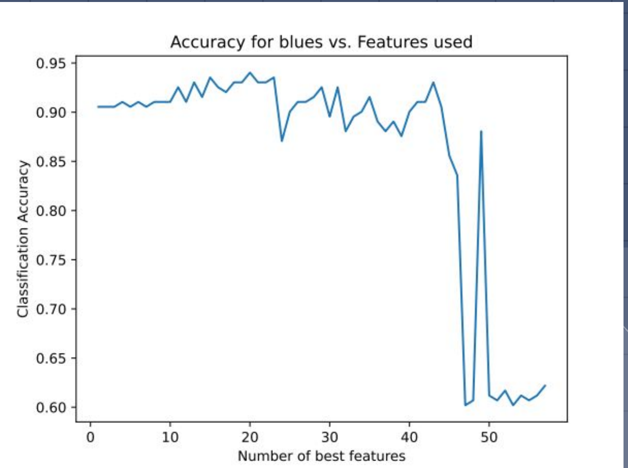
$$w \leftarrow w - a\nabla J_{xi}(w)$$

- Individual approach: Is this example genre y (yes or no)?

- Committee approach: Which genre is this example?

Logistic Regression Results





- Not a great predictor for blues
- Sacrifice true negatives for true positives

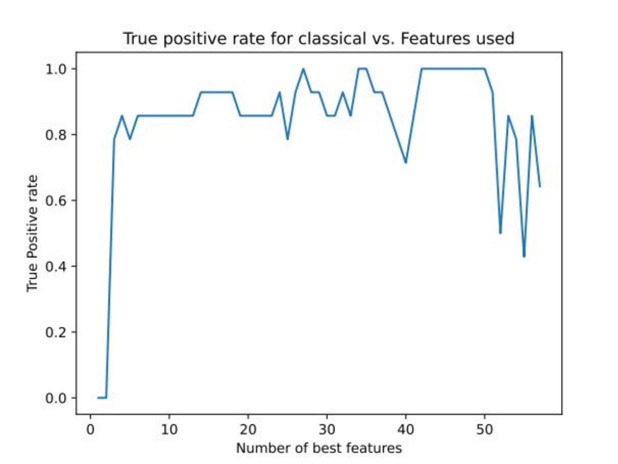
40 best features:

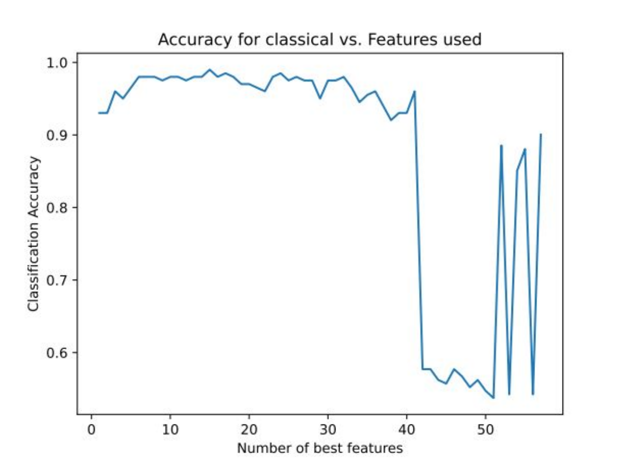
	not blues	blues
not blues	172	10
blues	10	9

All features:

	not blues	blues
not blues	108	74
blues	2	17

Logistic Regression results





- A great predictor for classical when using the top 30 best features
- Doesn't have the same problems as the blues model

All features:

	not classical	classical
not classical	172	15
classical	5	9

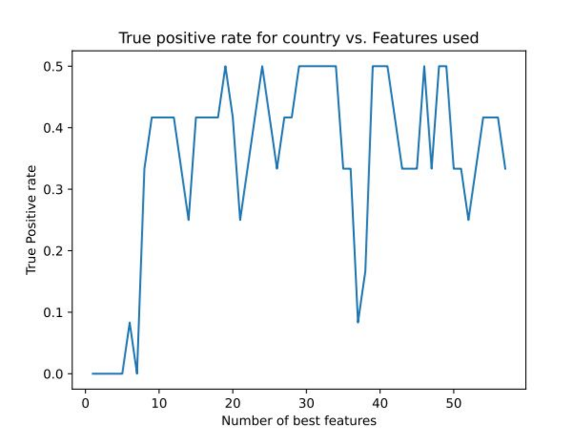
30 best features:

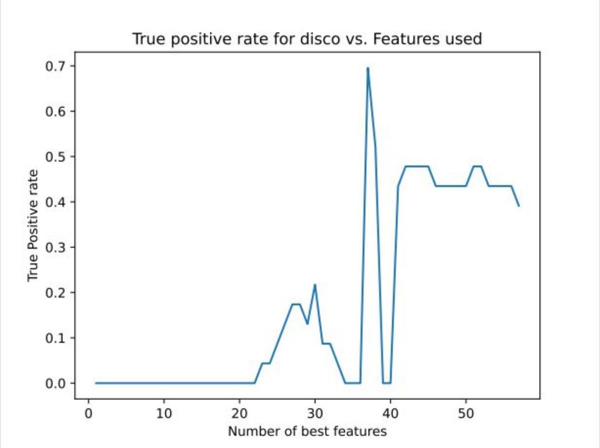
	not classical	classical
not classical	184	3
classical	2	12

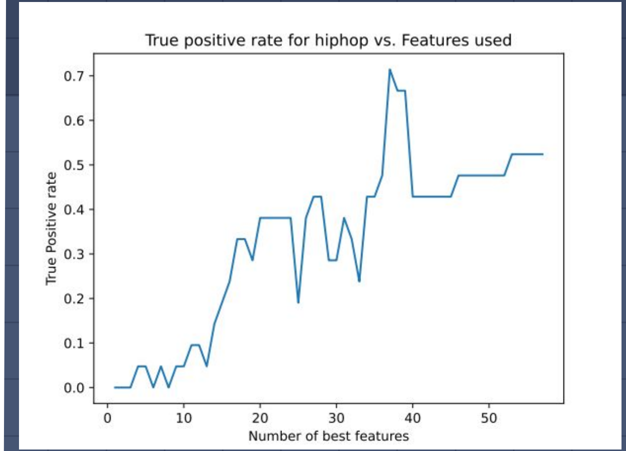
30 best features: (n=2000)

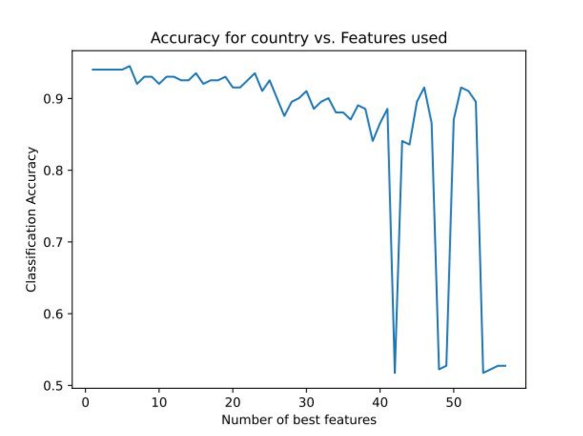
	not classical	classical
not classical	1779	22
classical	22	168

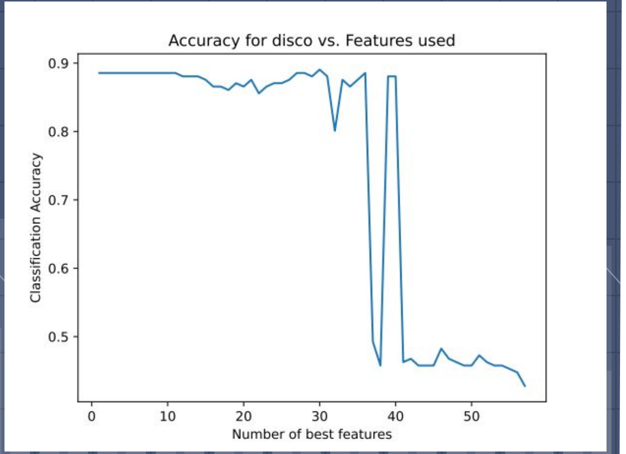
Logistic Regression results

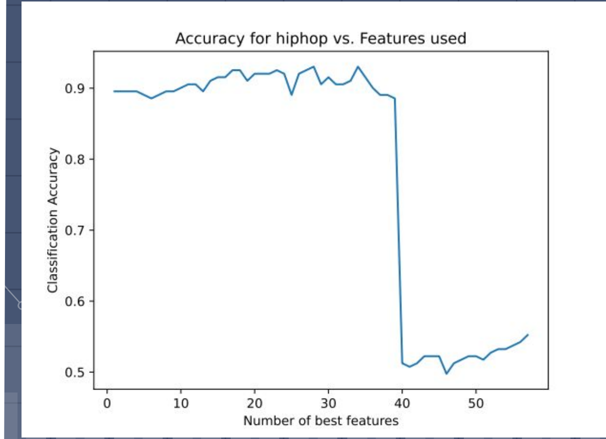




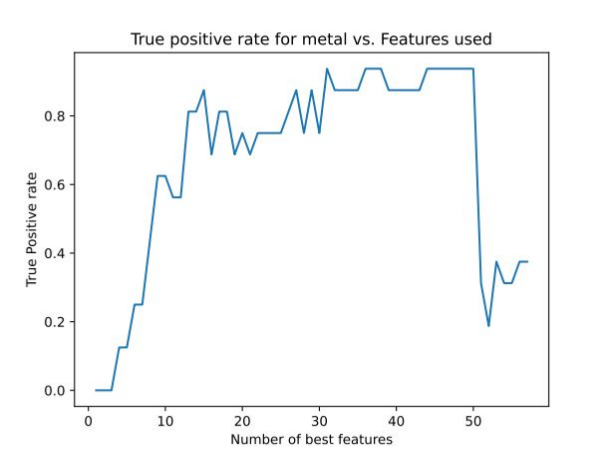


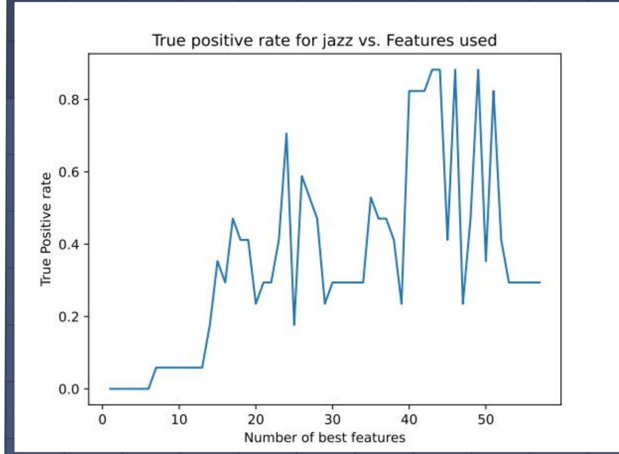


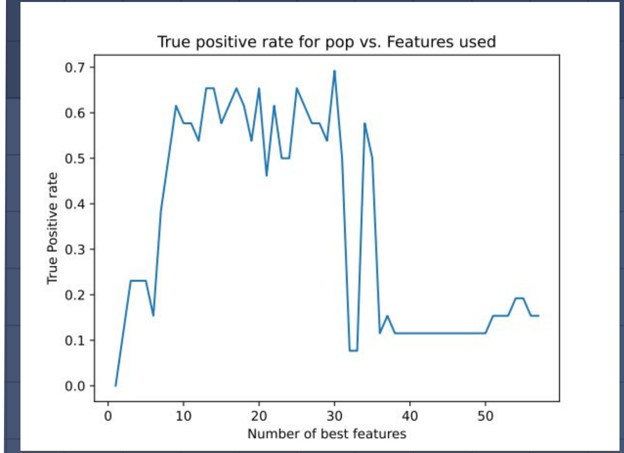


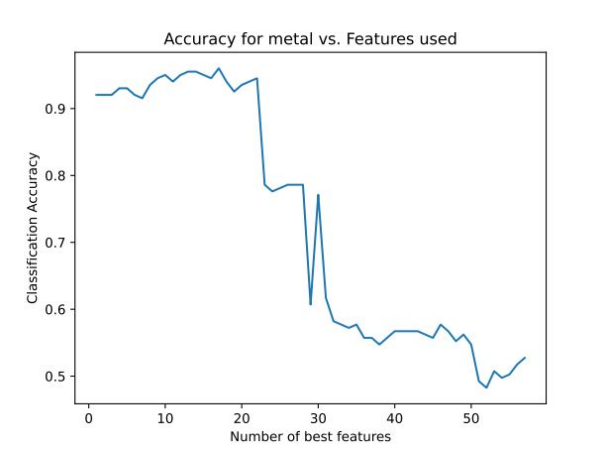


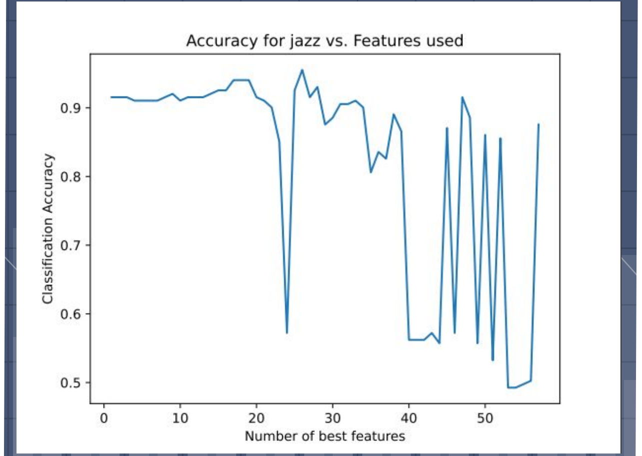
Logistic Regression results

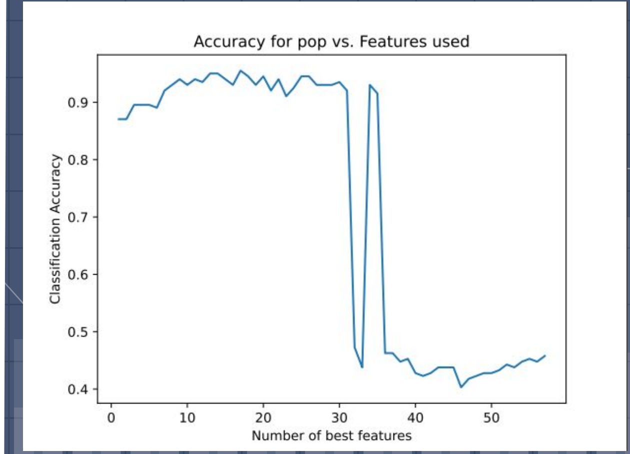




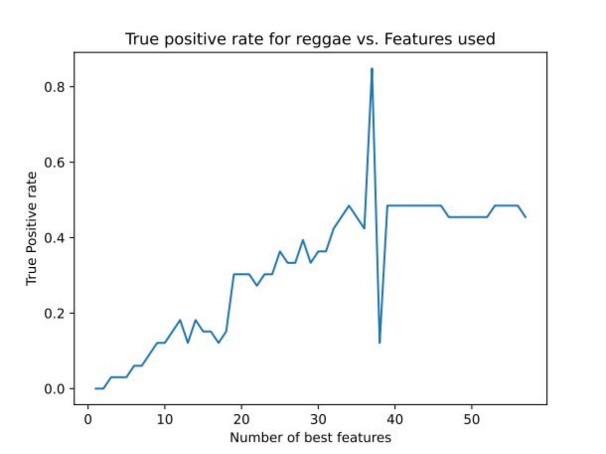


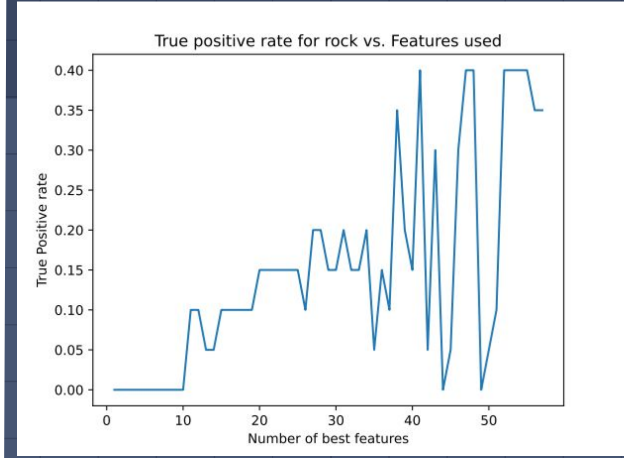


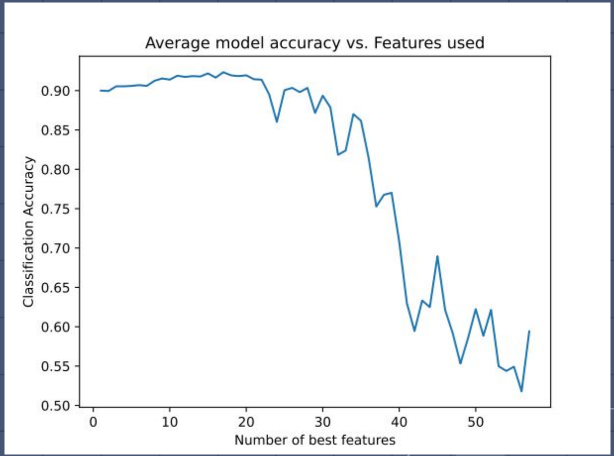


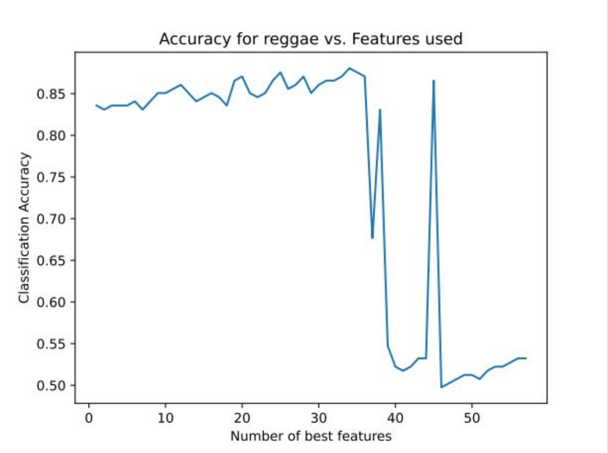


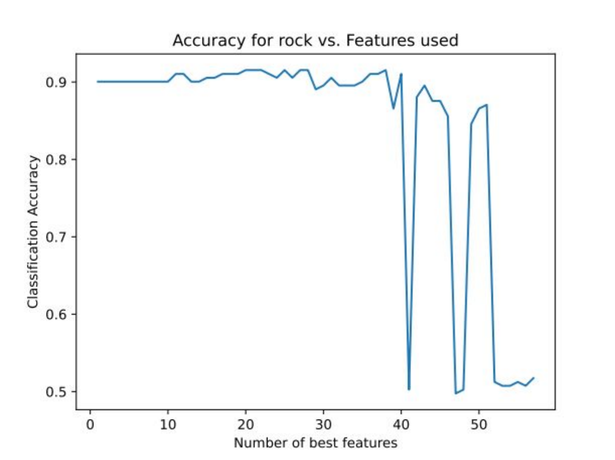
Logistic Regression results



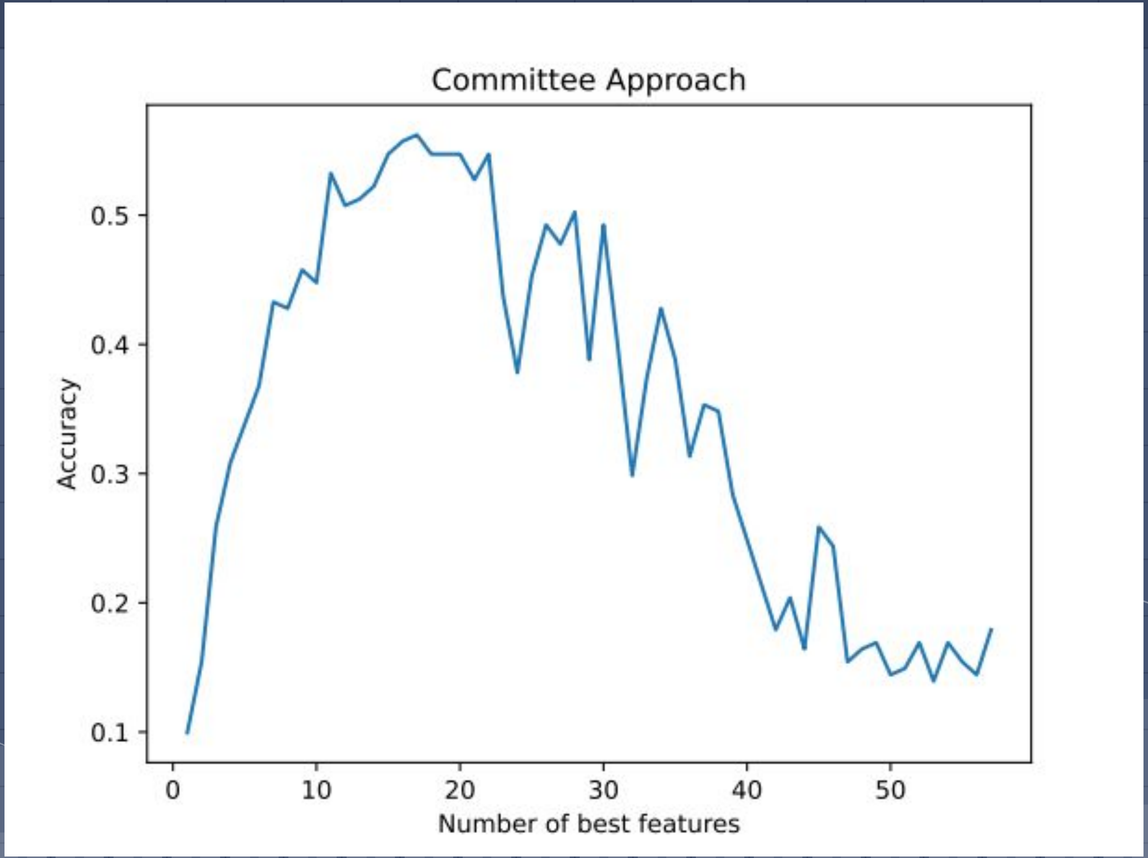








The Committee Method



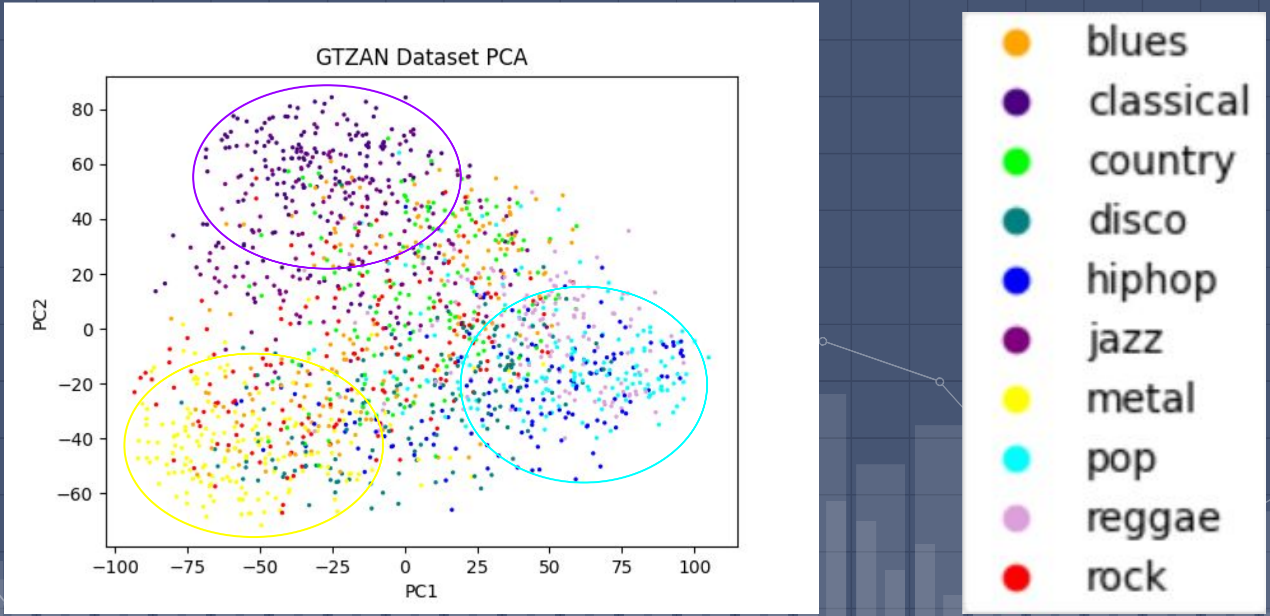
- ŷ=max of individual models
- Bad individual models limit performance
- Interestingly consistent with the average accuracy of the individual models

	blues	classical	country	disco	hiphop	jazz	metal	pop	reggae	rock
blues	13	0	1	0	0	1	4	0	0	0
classical	0	13	1	0	0	0	0	0	0	0
country	3	0	6	0	0	2	0	0	1	0
disco	0	0	3	3	4	0	5	3	3	2
hiphop	0	0	0	1	11	0	3	1	4	1
jazz	3	3	1	0	0	8	0	2	0	0
metal	0	0	0	1	0	1	14	0	0	0
pop	0	0	1	3	3	0	1	17	1	0
reggae	4	0	3	1	4	2	2	0	17	0
rock	1	1	3	1	0	3	3	0	0	8

accuracy: 0.5472636815920398

PCA Comparison

The best individual predictors match up with the most distinctive clusters in the PCA:



Conclusion, challenges, and limitations

- It is possible to predict some genres well, but not others
- Doing Naive bayes and SGD on large datasets many times is very resource and time consuming
- Working around this resource and time consumption was a challenge