Support Vector Musicality

What kinds of music do you listen to?
What do you like about it?

Any genres you *don't* like? Why?

12th ISMIR Conference in **2011**

release of the Million Song Database

- 280 GB of data
- 1,000,000 songs/files

- before the MSD…
  - 698 – 3227 songs
  - GTZAN, ISMIRGenre, ISMIRRhythm, Latin Music Database
- Collaboration between Columbia EE lab and the Echo Nest in MA *[1]*

- Acoustic Features
  - loudness
  - pitches
  - timbre
  - bars, beats, key, mode, tempo…
- Ground truth labels:
  - Avant Garde, Blues, Children, Classical, Comedy Spoken, Country, Easy Listening, Electronic, Folk, Holiday, International, Jazz, Latin, New Age, Pop Rock, Rap, Reggae, Religious, RnB, Stage, Vocal (`n_classes = 21`)

13th ISMIR Conference in **2012**

10th AMR Conference in **2012**

existing research literature on MSD

- Benchmark performances on various features & models
- SSD (Statistical Spectrum Descriptor)
  - SVM, *27.41%*
  - kNN, *27.07%*
- MFCC (timbral feature)
  - kNN, *24.13%*

- Schindler, Mayer, Rauber from Vienna University of Technology *[2]*

- Temporal features boost classification performance
- SSD for *each* segment + timbre & chroma features
  - SVM, *76.1%*
  - kNN, *68.1%*

- Schindler & Rauber from Vienna University of Technology *[3]*

considerations in scope & approach

- Limitations
  - Inability to **parallelize**
  - 7 weeks: proposal to poster

- Options
  - MIR is a very active field…
  - Apply novel features or models?
  - Minimize train & test errors?

Deliberated objective:

1. Rudimentary approximation of known optimal solution
   - SSD + timbre & chroma
   - SVM with polynomial kernel

2. Interpretation and analysis of fitted SVM parameters
   - Support vectors?
   - Dual coefficients?

Is music mathematical?? Or is math musical???

- Sonograms: visualizing sound

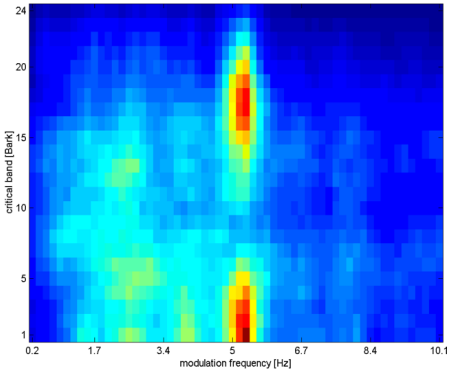


Figure 1a: rock music Sonogram

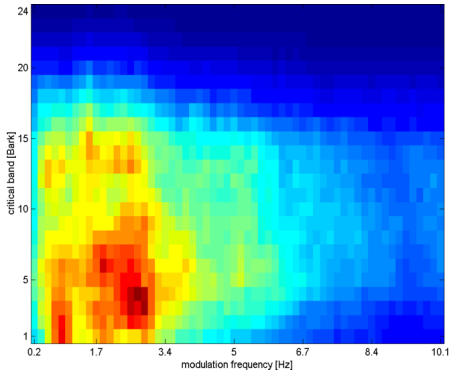


Figure 1b: classical music Sonogram

- **Amplitudes** over 60 **frequencies** (grouped into 24 'critical bands'), reflects 'loudness' of different pitches perceived by humans

- Statistical Spectrum Descriptor
  - *For each critical band*, compute statistical moments on Sonogram:
  - mean, median, variance, skewness, min- and max-value

- **Timbre:** tone color/quality; distinguishes sound textures of *the same pitch and loudness* produced by different instruments
  - time-domain zero crossing
  - spectral centroid, rolloff, flux
  - Mel-Frequency Cepstral Coefficients

- **Chroma**: capture harmonic & melodic characteristics, 12 pitch classes from C, C#, D, … A#, B

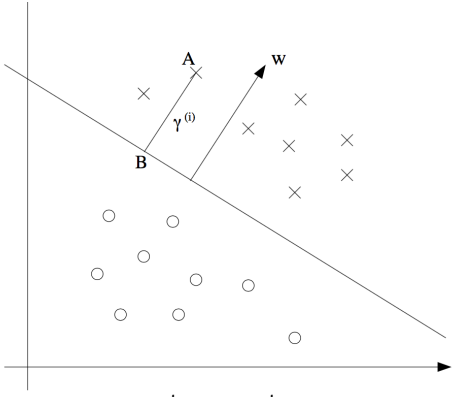
- Western music notations characterize a pitch by tone height and chroma

support vectors as avant-garde fusion genres??

`sklearn.svm.SVC`:

$$w^T x + b = \left(\sum_{i=1}^{m} \alpha_i y^{(i)} x^{(i)} \right)^T x + b$$
$$= \sum_{i=1}^{m} \alpha_i y^{(i)} \langle x^{(i)}, x \rangle + b.$$

- one-vs-rest approach
  - *`n_class`* total number of classifiers & hyperplanes
- `support_vectors_`
  - songs closest to separating hyperplane
  - *least* canonical songs of certain genres



- `dual_coef_`
  - features with largest magnitude for the corresponding support vector
    ➔ most defining sound frequencies that distinguish/identifies the genre
  - dimensionality (`n_class – 1) x n_SV`

experimental results

- Classifier
  - Support Vector Machine
  - Kernel: 3rd degree polynomial
  - C = 10
  - one-vs-rest multi-class approach

Train accuracy	92.45%
Test accuracy	35.80%

- Dataset
  - 292 columns
    - 168: SSD for 24 critical bands
    - 124: timbre and chroma
  - Train/test split
    - Trained with 2,000 samples per class
    - Test sample distribution:

Electronic	38,540
International	12,181
Jazz	15,741
Latin	15,475
Pop Rock	235,214
Rap	18,855
RnB	12,304

Comedy Spoken	*56*
Country	9,686
Folk	3,784
New Age	*1,992*
Reggae	4,898
Religious	6,775
Vocal	4,179

	blues	comedy	country	electr	folk	intnl.	jazz	latin	new age	rock	rap	reggae	relig	r n b	vocal
blues	1869	5	4	3	14	23	20	18	9	5	3	6	5	4	12
comedy	15	1883	14	6	14	14	3	12	1	1	0	4	7	10	16
country	19	12	1839	2	25	19	10	17	8	2	1	3	18	8	17
electronic	10	2	7	1871	9	32	6	15	4	4	6	13	5	13	3
folk	30	7	45	3	1780	35	8	28	15	2	0	6	14	6	21
International	18	3	16	4	31	1821	11	24	13	3	11	16	5	6	18
jazz	22	3	10	5	12	31	1814	27	21	3	0	12	8	16	16
latin	25	2	23	4	28	29	10	1821	6	4	5	9	6	14	14
new age	13	1	14	12	22	50	19	22	1821	2	0	0	7	14	3
pop rock	32	3	33	16	35	45	14	34	8	1719	3	12	16	22	8
rap	4	7	6	8	4	24	4	14	1	2	1869	38	2	16	1
reggae	14	1	6	2	2	21	7	25	0	1	7	1902	2	8	2
religious	20	5	36	1	25	38	7	37	7	6	7	8	1767	24	12
r n b	29	1	20	6	23	26	10	26	9	4	6	20	8	1799	13
vocal	32	4	35	3	27	30	35	18	10	3	2	5	8	18	1770
misclassified	283	56	269	75	271	417	164	317	112	42	51	152	111	179	156
n_support	1942	1244	1893	1733	1975	1998	1969	1970	1775	1795	1634	1793	1932	1919	1930

future work

- Hardware and software infrastructures for **parallel** data processing
- Robust **debugging** for Python and ML libraries
- *Refuse to succumb to The Black Box!!!*
- NP-hard: many approaches to column selection…

references

[1] D. P. W. E. Thierry Bertin-Mahieux and P. L. Brian Whitman, “The million song dataset,” in Proceedings of the 12th International Conference on Music Information Retrieval, 2011.

[2] R. M. Alexander Schindler and Andreas Rauber, “Facilitating comprehensive benchmarking experiments on the million song dataset,” in Proceedings of the 13th International Society for Music Information Retrieval Converence, 2012.

[3] A. R. Alexander Schindler, “Capturing the temporal domain in Echonest Features for improved classification effiectiveness,” in Proceedings of the 10th International Workshop on Adaptive Multimedia Retrieval, 2012.