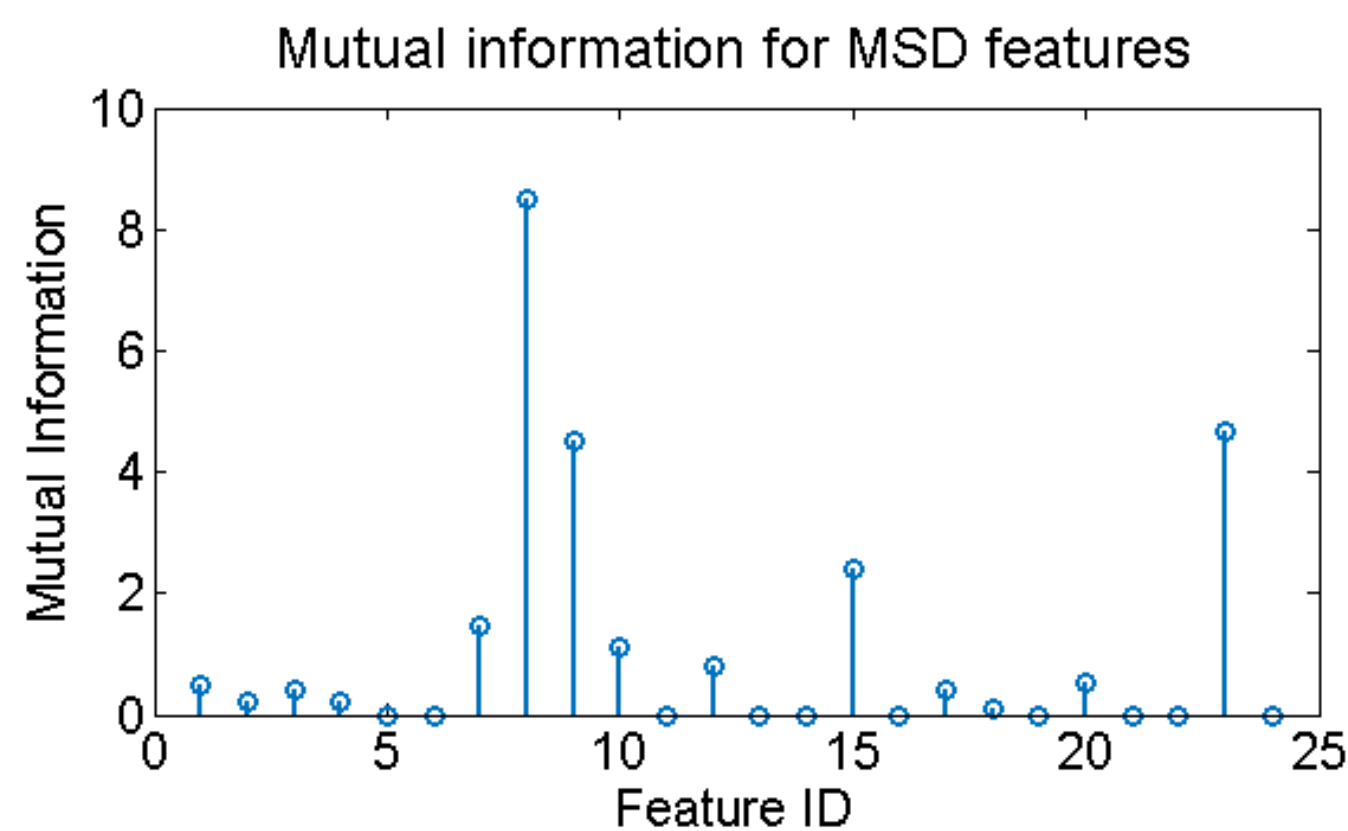
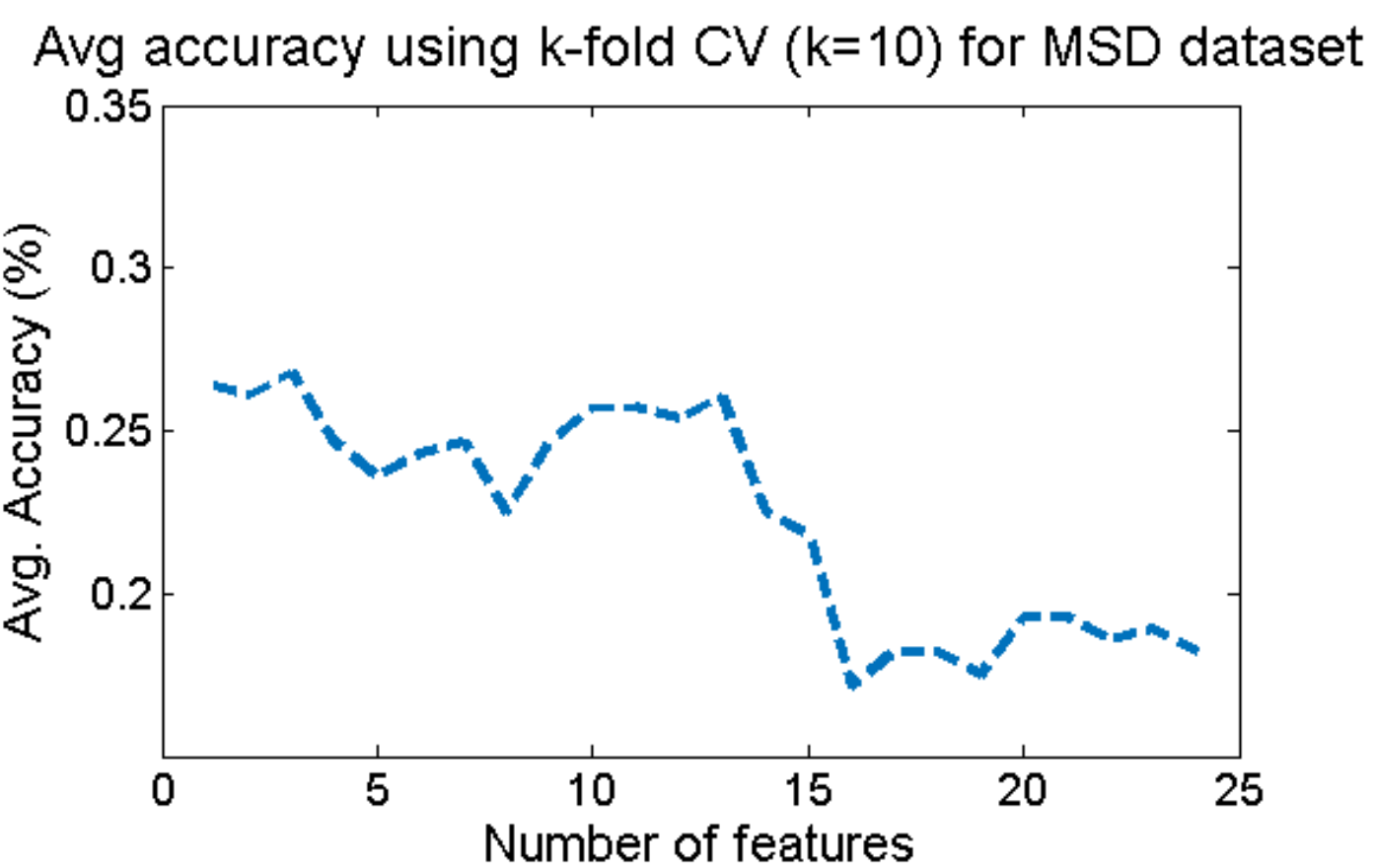




### Data sources

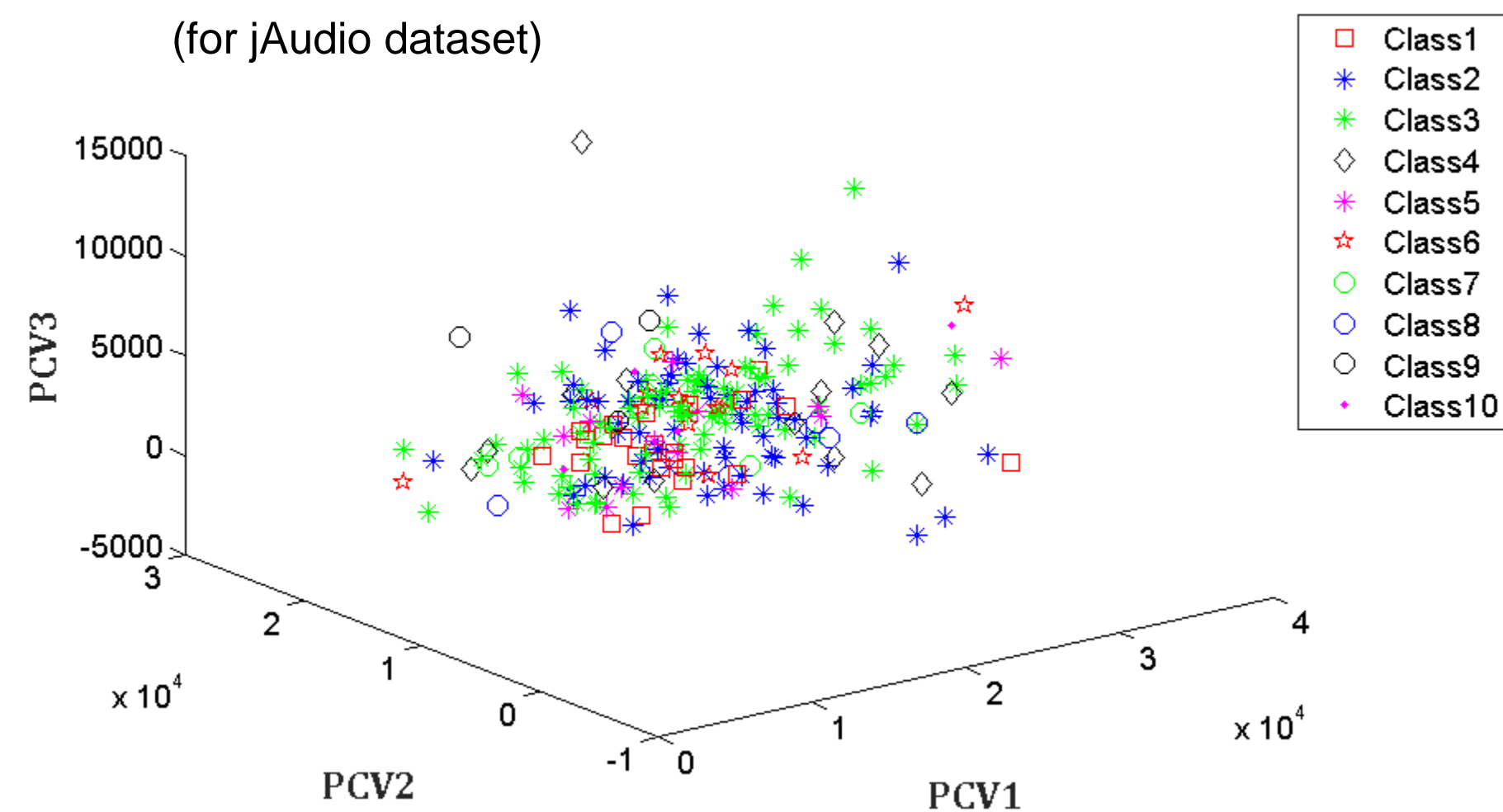
Feature	Workspace
<b>Strongest Beat</b>	Name
Derivative of Strongest Beat	abc F10_artist_location
Running Mean of Strongest Beat	abc F11_artist_name
Standard Deviation of Strongest Beat	abc F12_release
Derivative of Running Mean of Strongest Beat	abc F13_release_7digid
Derivative of Standard Deviation of Strongest Beat	abc F14_song_id
<b>Beat Sum</b>	abc F15_song_hottness
Derivative of Beat Sum	abc F16_title
Running Mean of Beat Sum	abc F17_track_7digid
Standard Deviation of Beat Sum	abc F18_similar_artists
Derivative of Running Mean of Beat Sum	{ F19_artist_terms
Derivative of Standard Deviation of Beat Sum	{ F1_num_songs
<b>Strength Of Strongest Beat</b>	{ F20_artist_terms_freq
Derivative of Strength Of Strongest Beat	{ F21_artist_terms_weight
Running Mean of Strength Of Strongest Beat	abc F22_analysis_sample_rate
Standard Deviation of Strength Of Strongest Beat	abc F23_audio_md5
Derivative of Running Mean of Strength Of Strongest Beat	abc F24_danceability
Derivative of Standard Deviation of Strength Of Strongest Beat	abc F25_duration
<b>Strongest Frequency Via Zero Crossings</b>	abc F26_end_of_fade_in
	abc F27_energy

### Feature extraction

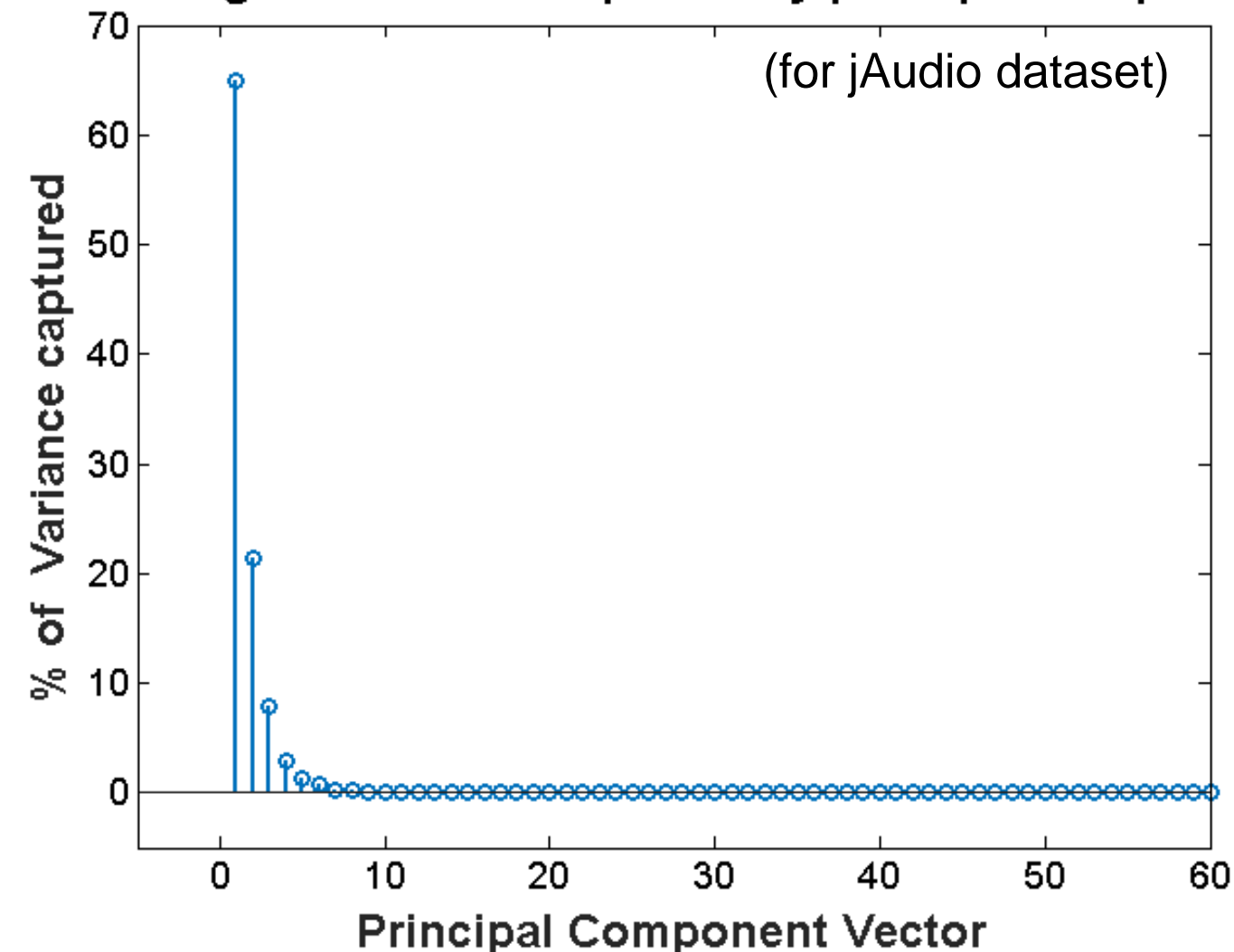


### Data exploration

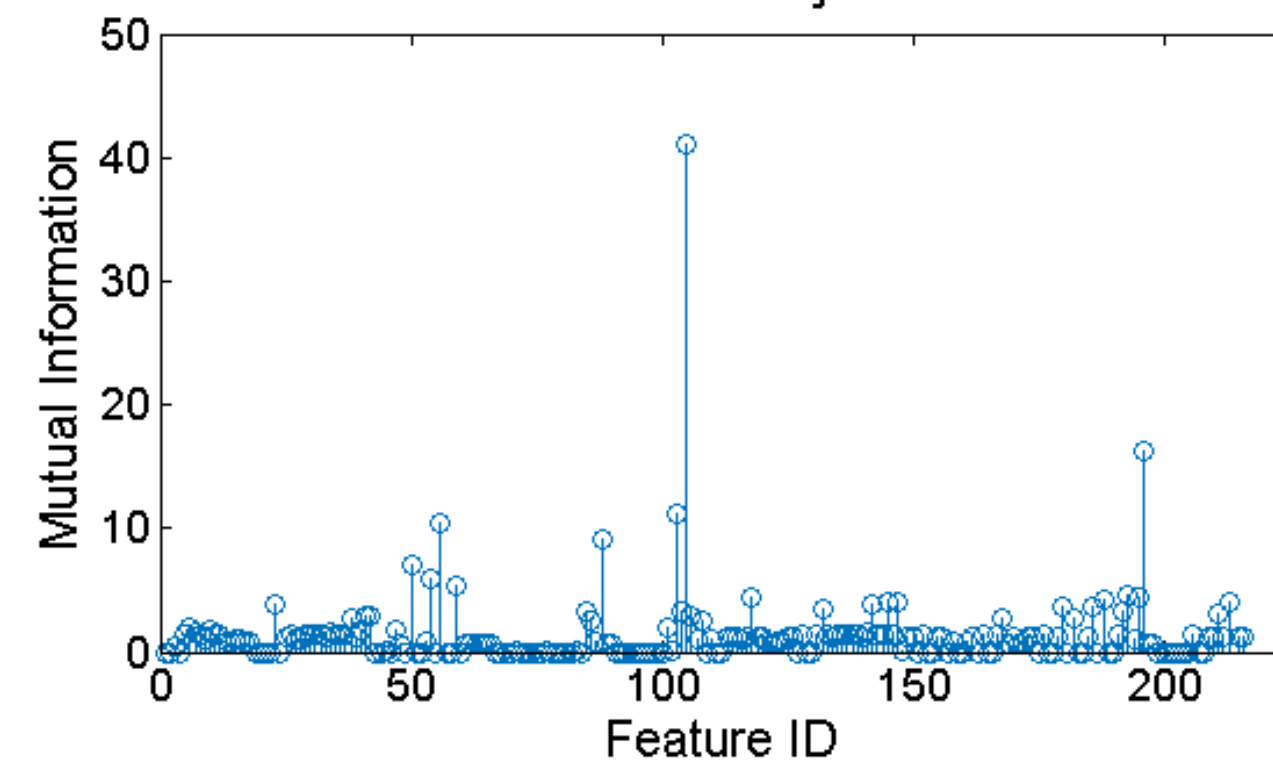
Data Projected onto three dimensions using PCA  
(for jAudio dataset)



Percentage of variance captured by principal components



Mutual information for jAudio features



### Algorithms

#### Prediction Accuracy:

Algorithm	MSD dataset	jAudio dataset
GDA	33.52%	29.03%
GDA modified	32.8%	19.31%
SVM	30%	40.28%

#### Error (5 classes) with MSD dataset:

Error span	GDA	GDA modified	SVM
1 class	45.82%	44.86%	42%
2 class	11.92%	12.3%	17.7%
3 class	8.58%	9.28%	7.4%
4 class	0.16%	0.76%	4.28%

#### Error (5 classes) with jAudio dataset:

Error span	GDA	GDA modified	SVM
1 class	25.28%	43.19%	15.28%
2 class	24.58%	25.83%	29.17%
3 class	20.28%	11.53%	13.89%
4 class	0.01%	0.001%	1.39%

### References

- [1] McEnnis, Daniel, Cory McKay, Ichiro Fujinaga. 2006. "Overview of On-demand metadata extraction network (OMEN)". ISMIR.
- [2] Thierry Bertin-Mahieux, Daniel P.W. Ellis, Brian Whitman, and Paul Lamere. "The Million Song Dataset". In Proceedings of the 12th International Society for Music Information Retrieval Conference (ISMIR), 2011
- [3] Kevin Wang, 2014 "Predicting hit songs with MIDI Musical Features", CS229 Project – 2014.