MUSIC GENRE SPOTIFICATION

Abhijit Suresh
Paria Rezaeinia
Sahana Sadagopan

PROBLEM STATEMENT

- Classify music into its corresponding genres.
- Not straightforward to compare distance between two songs
- Curse of dimensionality.

Classical: 320 songs

Electronics: 115 songs

Jazz/blues: 26 songs

Metal/punk: 45 songs

Rock/pop: 101 songs

World: 122 songs

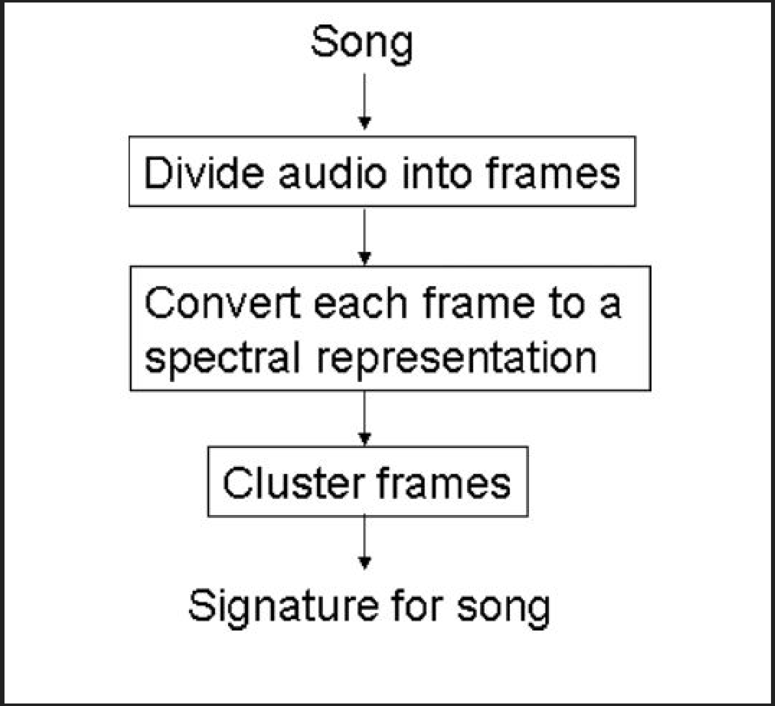
OVERVIEW OF OUR APPROACH

- Dimensionality reduction
  - Mel Frequency Cepstrum Coefficients (MFCC)
  - Principal Component Analysis (PCA)
  - k-Means
  - Multidimensional scaling
  - Gaussian Mixture (Modified)

- Distance calculation
  - Minowski distance
  - Earth Movers distance
  - Euclidean distance
  - Kullback-Leibler distance (KL) distance

- Classification
  - Neural Network
  - k- Nearest Neighbours (kNN)

Approach 1: Content based similarity method



Johnson lindenstrauss n > 77 (79)

Euclidean distance

$$P = \{(\mu_{p_1}, \Sigma_{p_1}, w_{p_1}), \ldots, (\mu_{p_m}, \Sigma_{p_m}, w_{p_m})\}$$

A Content-Based Music Similarity Function

Beth Logan   *Ariel Salomon*

Approach 1: Earth Mover's distance

- The earth mover's distance (EMD) is a measure of the distance between two probability distributions over a region D

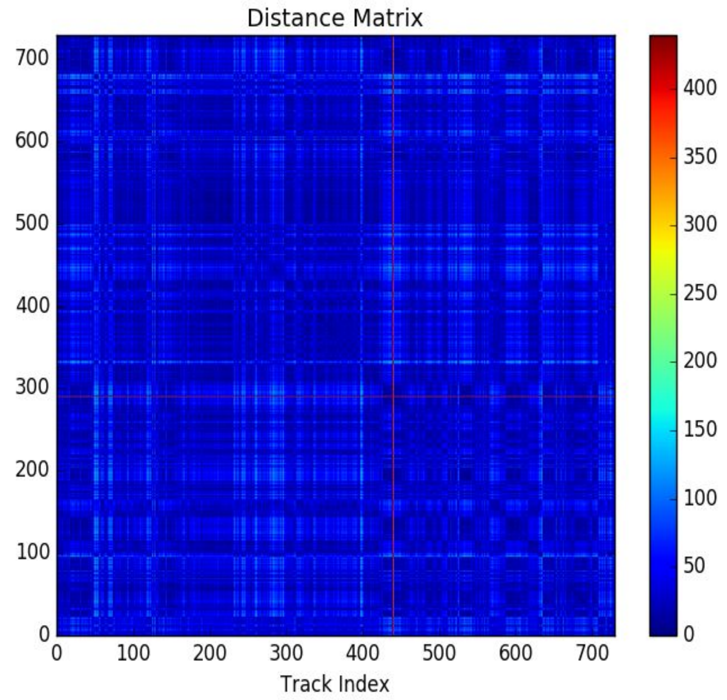
- It is also known as the Wasserstein metric.

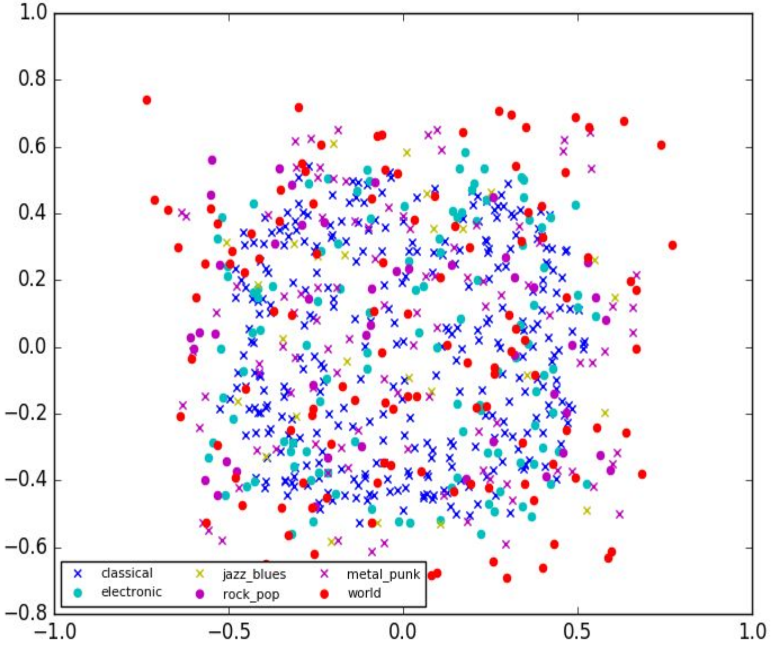
$$EMD(P,Q) = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{i,j} d_{i,j}}{\sum_{i=1}^{m} \sum_{j=1}^{n} f_{i,j}}$$

$$\sum_{i=1}^{m} \sum_{j=1}^{n} f_{i,j} = \min \left(\sum_{i=1}^{m} w_{pi}, \sum_{j=1}^{n} w_{qj} \right)$$

$D = [d_{i,j}]$ be the ground distance between clusters p_i and q_j.

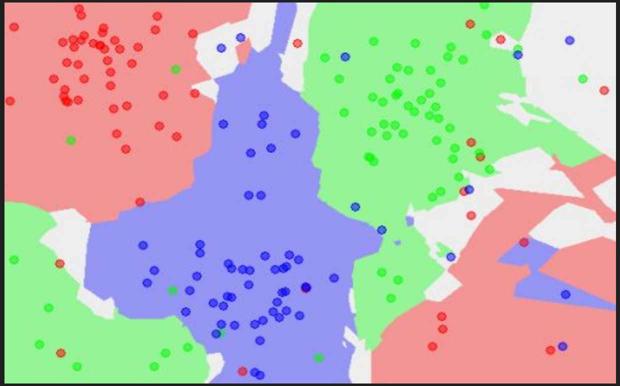
Approach 1: Multidimensional scaling

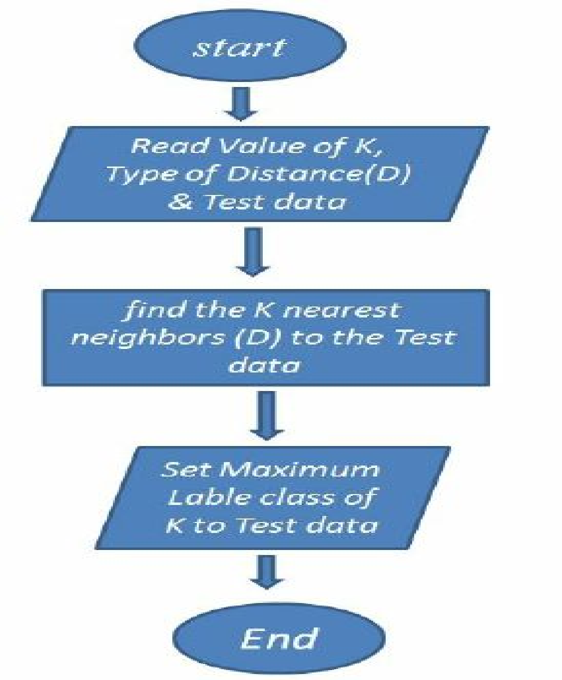




Classification of Approach 1: kNN

- The Distance vectors are taken which are in metric space
- Supervised Method of learning algorithm
- sorts the distance and determine k nearest neighbors based on the k-th minimum distance
- Accuracy: 54%





Cross Validation

We have used 10-cross validation for our results

- 90% of the data was taken as training data and 10% was taken as test data in case of KNN classification method

This was performed 10 times to predict the efficiency.

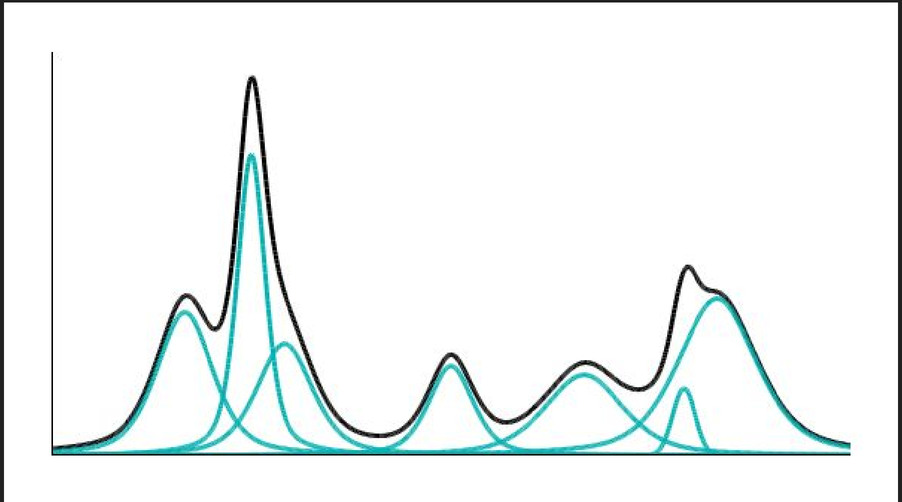
Classification of Approach 1: Confusion matrix (n = 20)

Accuracy = 54%

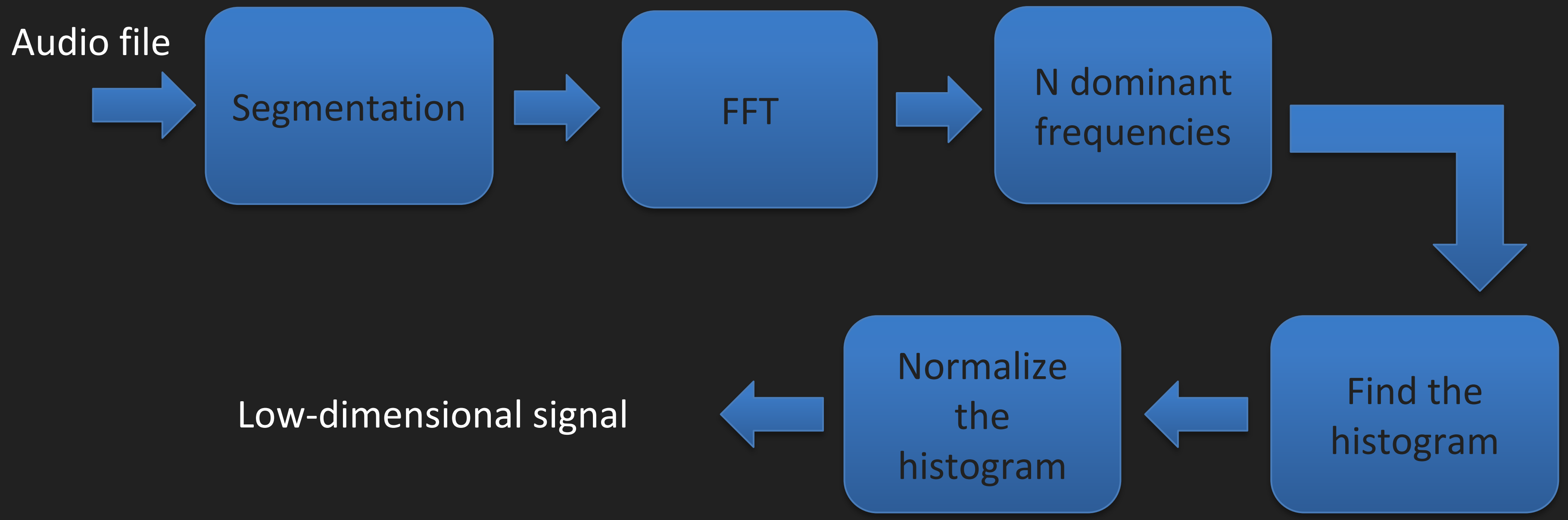
	Classical	**Electronic**	**Jazz-blues**	**Metal-punk**	**Rock-pop**	**world**
Classical	**0.6491**	0.1403	0.017	0.035	0.035	0.122
Electronic	0.3333	**0.6666**	0	0	0	0
Jazz-blues	0	0	**0**	0	0	0
Metal-punk	0	0	0	**0**	0	0
Rock-pop	0	0.2	0.2	0	**0**	0.6
World	0.125	0.125	0	.0	0.25	**0.5**

Gaussian Mixture Model

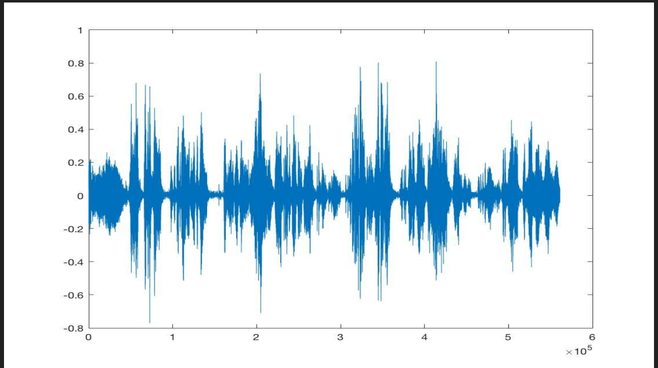
- A parametric probability density function
- A weighted sum of Gaussian components density
- Individual component densities model the underlying set of hidden classes



Approach 2: FFT Distance Matrix

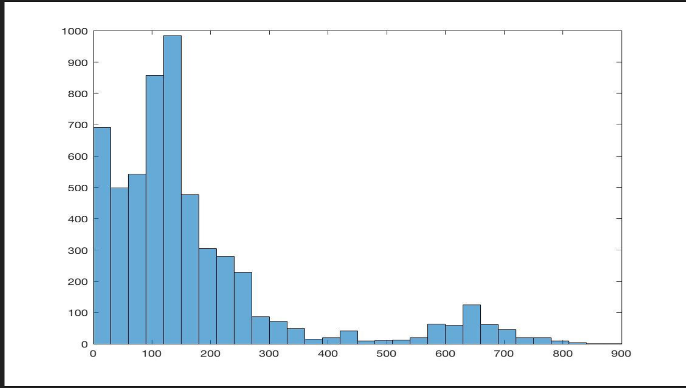


Approach 2: Example



Audio file

FFT
Distance
Matrix
Algorithm



Distribution

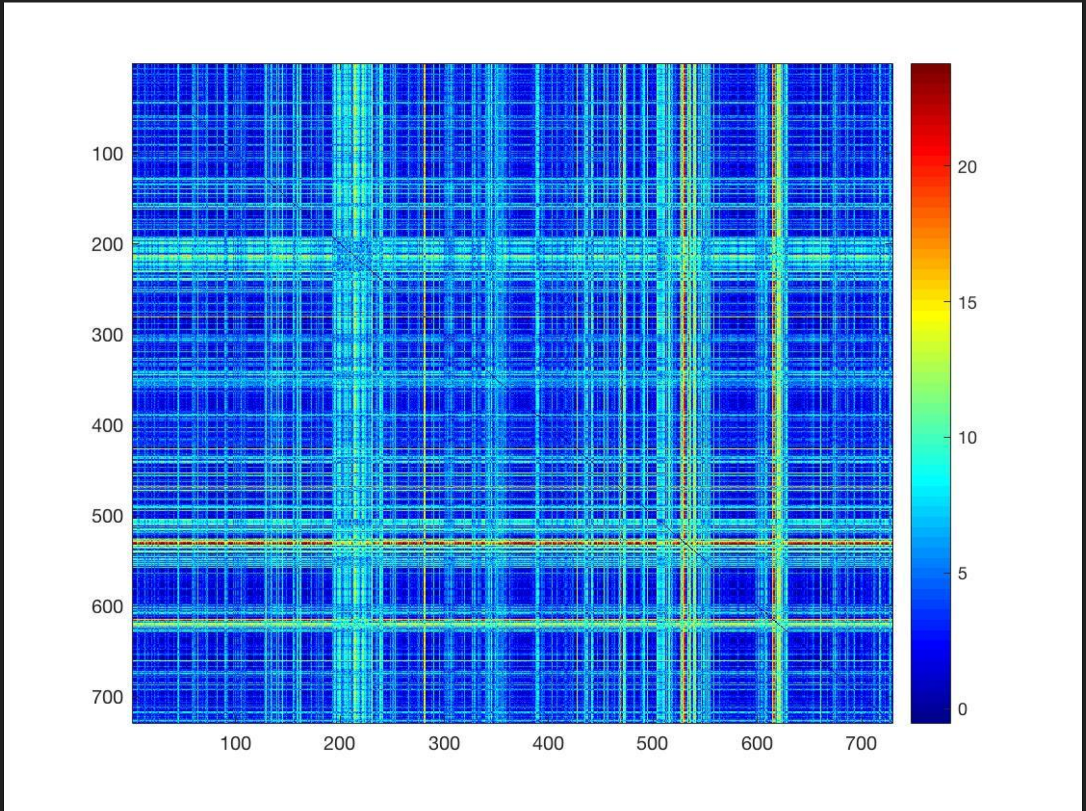
Kullback-Leibler divergence metric

- A measure of distance between two probability distributions

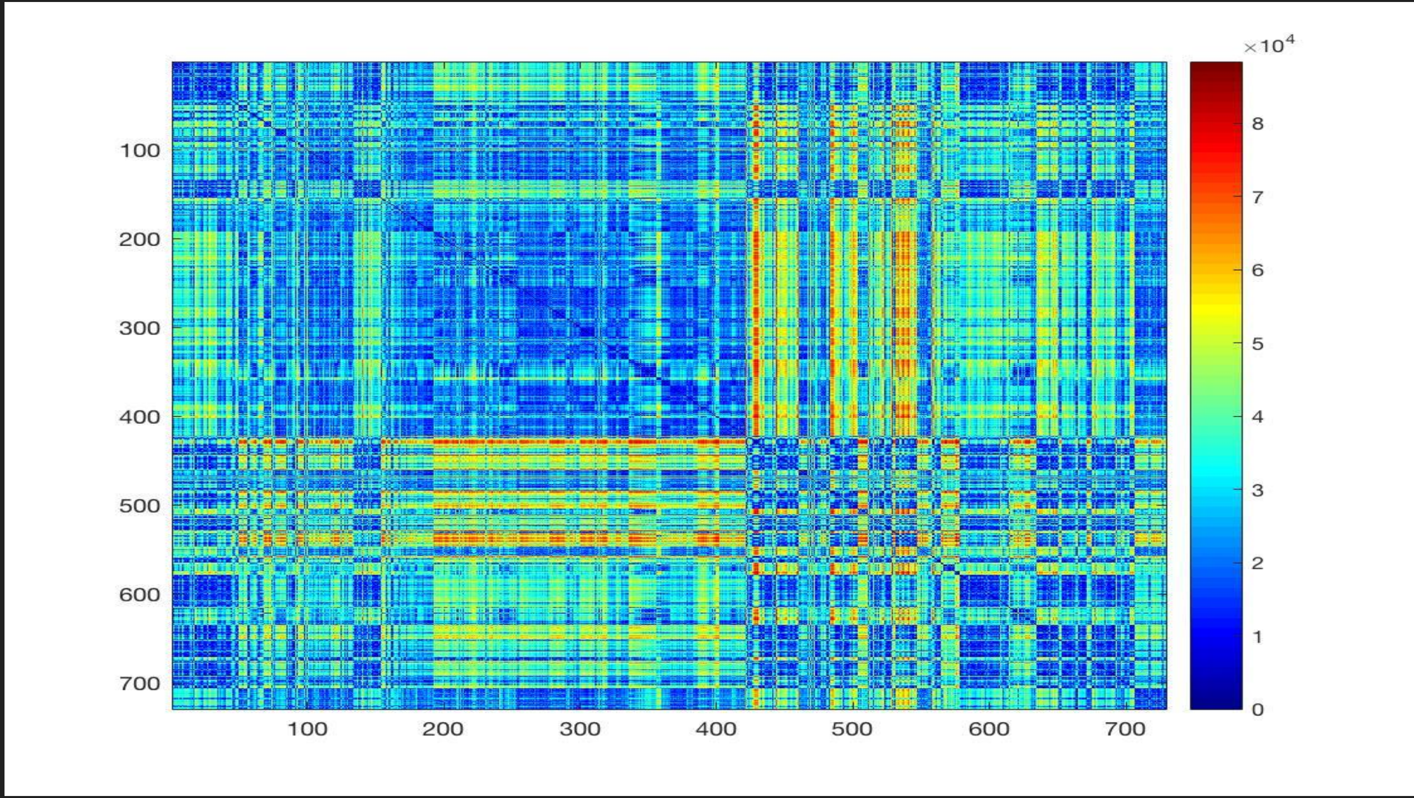
  $D(P||Q) = H(P,Q) - H(P)$

- Not symmetric

Distance Matrix



Approach 2 : Distance matrix



Confusion matrix (d = 30)

Accuracy = 52%

	Classical	**Electronic**	**Jazz-blues**	**Metal-punk**	**Rock-pop**	**world**
Classical	.8329	.0919	0	0	.0111	.0641
Electronic	.5182	.3636	0	0	.1	.0182
Jazz-blues	.4231	0	0	.0769	.3077	.1923
Metal-punk	.4286	.0612	0	.0612	.1224	.3265
Rock-pop	.5439	.0526	0	.0175	.2281	.1579
world	.6522	.0543	.0109	.0109	.0217	.25

Confusion matrix (d = 50)

Accuracy = 57%

	Classical	Electronic	Jazz-blues	Metal-punk	Rock-pop	world
Classical	.8319	.0551	0	0	.0377	.0754
Electronic	.6106	.2832	0	.265	.0619	.0177
Jazz-blues	.1304	0	0	.2609	.087	.5217
Metal-punk	.4750	.075	0	.125	.15	.175
Rock-pop	.5109	.0543	0	.0217	.2609	.1522
world	.5701	.0187	.0093	.0187	.0561	.3271

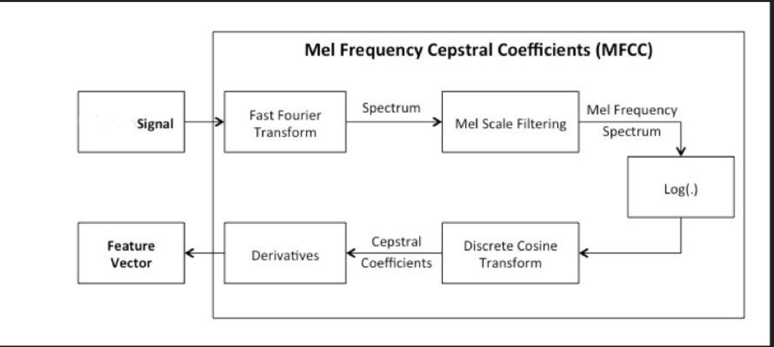
Confusion matrix (d = 50), modified kNN

Accuracy = 63%

	Classical	Electronic	Jazz-blues	Metal-punk	Rock-pop	world
Classical	.5821	.1422	.0228	.0998	.0717	.0814
Electronic	.0442	.7876	.0531	.0708	.0177	.0265
Jazz-blues	0	0	1	0	0	0
Metal-punk	0	0	0	.8448	.0345	.0345
Rock-pop	.0648	.0463	.1019	.1019	.6759	.0093
world	.0283	.0943	.1698	.1038	.1321	.4717

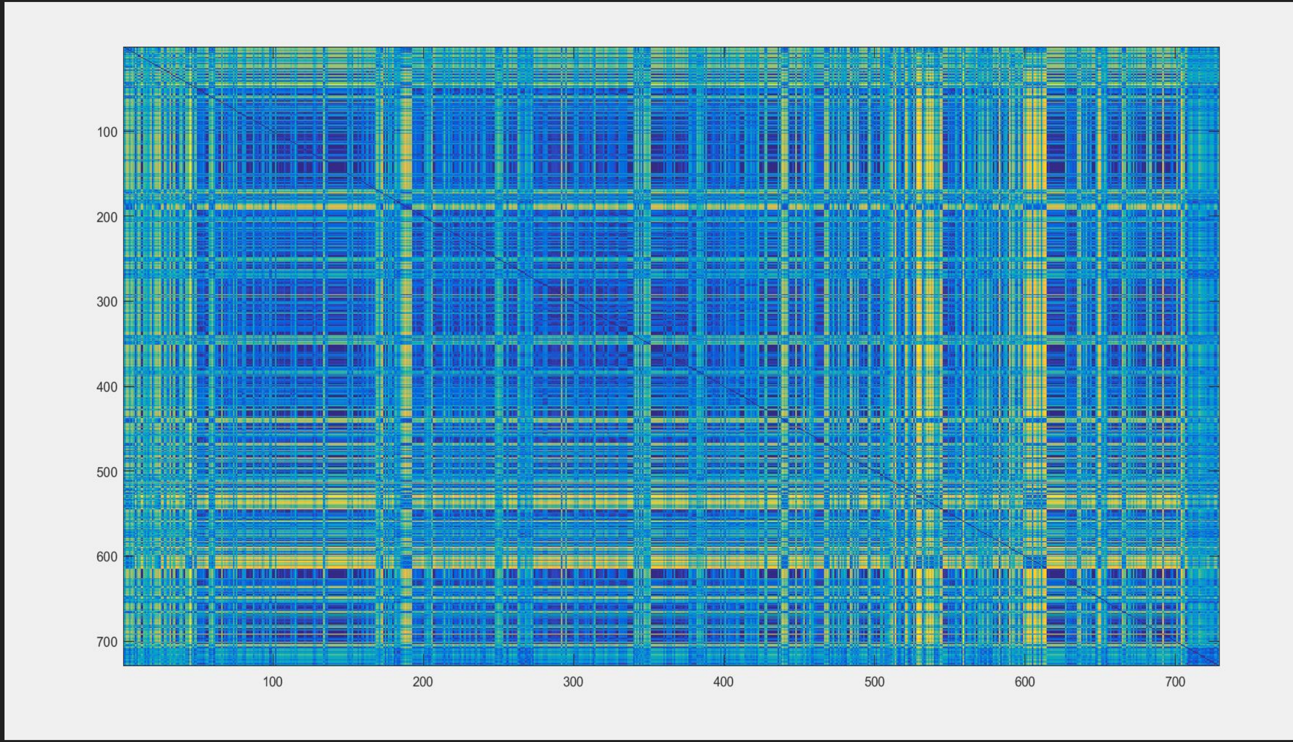
Approach 3 : MFCC and PCA

- MFCC or Mel Frequency Cepstrum Coefficients



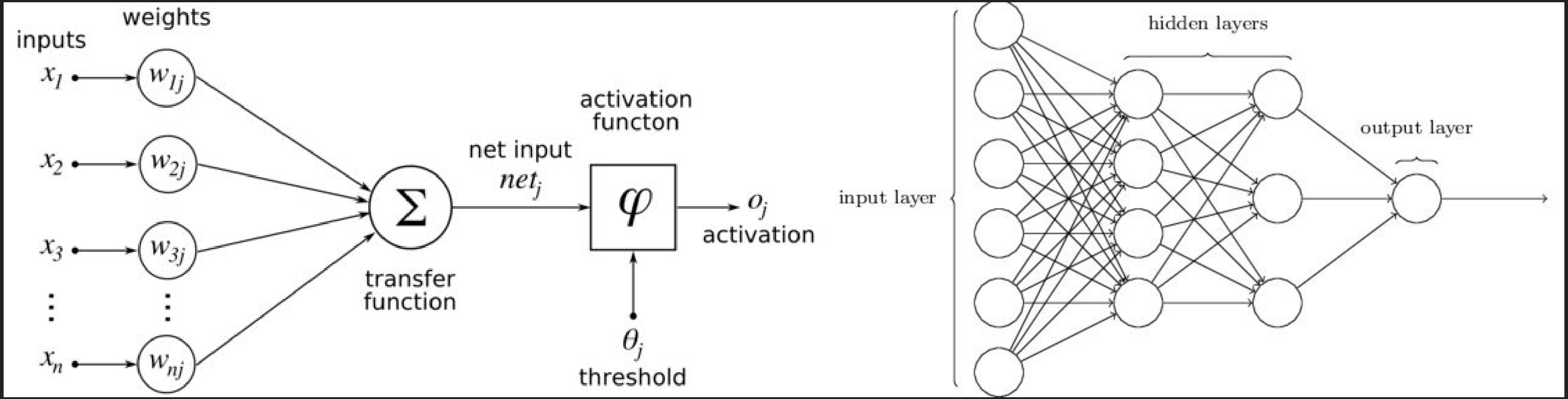
- PCA or Principal Component analysis
    - Identify principal components
    - The transformation is defined in such a way that the first principal component has the largest possible variance. This applies for the reset of components.

Approach 3 : Distance Matrix



Classification of Approach 3: NEURAL NETWORK

Forward propagation-backward propagation model of neural network

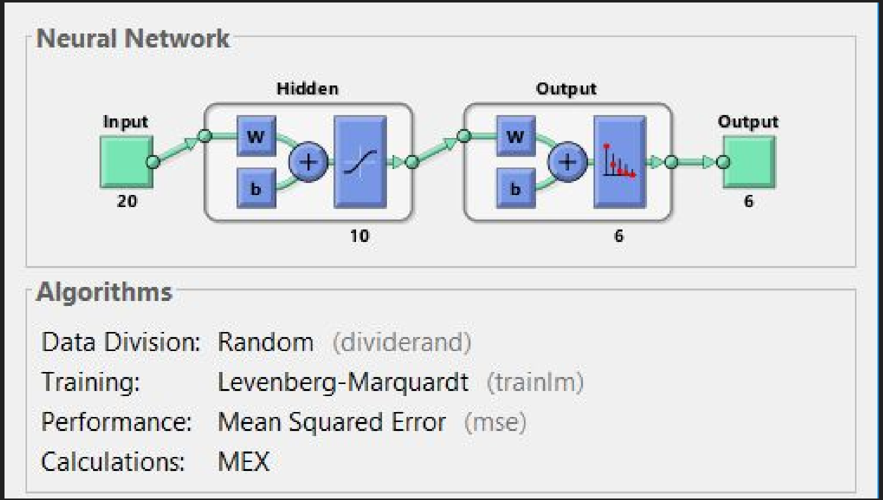


Neural network Training

The Input was taken as the Songs of all 6 genre and the Target data as the distance matrix for features.

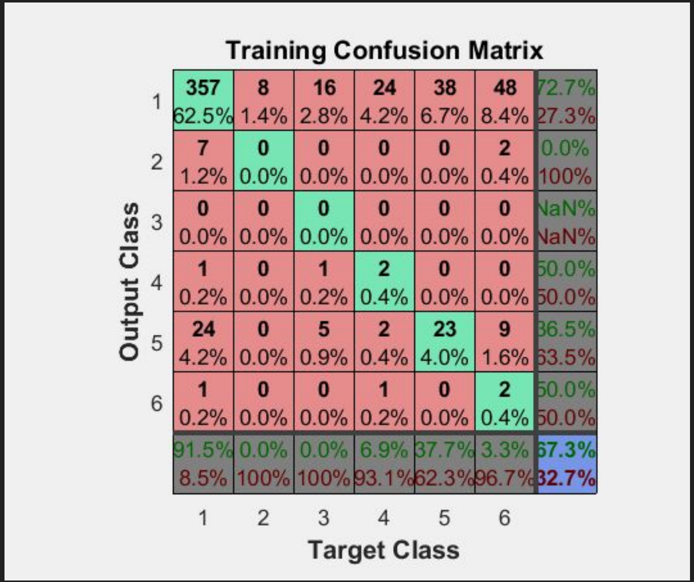
The Parameters that play a role in Neural Network

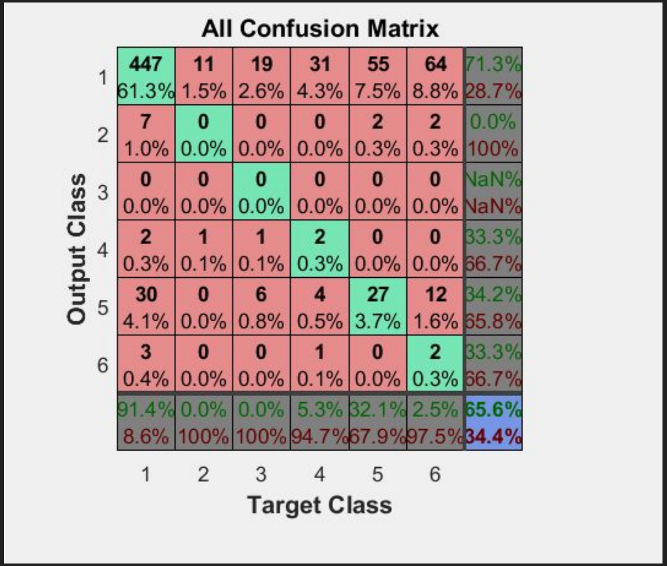
- The Interconnection Patterns
- Learning Process for updating weights
- Activation function



Classification of Approach 3: Results

The Confusion Matrix obtained For neural network method classification





Results

The Accuracy level for different approaches after cross validation are shown below:

Dimensionality Reduction Technique	Distance Calculation	Classification Method	Efficiency Obtained (After Cross Validation)
Content based similarity method	Earth Mover Distance	KNN algorithm	54%
Gaussian Mixture (Modified)	Euclidean Distance	KNN Algorithm	57%
Gaussian Mixture (Modified)	Euclidean Distance	Modified - KNN Algorithm	63%
MFCC & PCA	Euclidean Distance	Neural Network Algorithm	65%

SUMMARY

- Importance of Dimensionality reduction
- The genre that were easily classified

  Classical genre

  Classical Genre composes nearly 50% of the data set given

- The genre that were classified badly

  Jazz was the one classified the worse

  Only 3% of the data is Jazz

NEXT STEPS

- When Clustering frames, we could try using Kullback leibler (KL metric) instead of euclidean distance.
- Use particle swarm optimization in order to find the optimal set of synaptic weights for classification.
- Try the algorithm on a different dataset where the genre types are distributed uniformly.

THANK YOU

Francois Meyer

References:

1. Logan, B., & Salomon, A. (2001). A content-based music similarity function. Cambridge Research Labs-Tech Report.

2. Lee, C. H., Shih, J. L., Yu, K. M., & Lin, H. S. (2009). Automatic music genre classification based on modulation spectral analysis of spectral and cepstral features. IEEE Transactions on Multimedia, 11(4), 670-682.

3. Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. IEEE Transactions on speech and audio processing, 10(5), 293-302.

4. Pampalk, E. (2006). Computational models of music similarity and their application to music information retrieval (Doctoral thesis).Vienna University of Technology, Austria