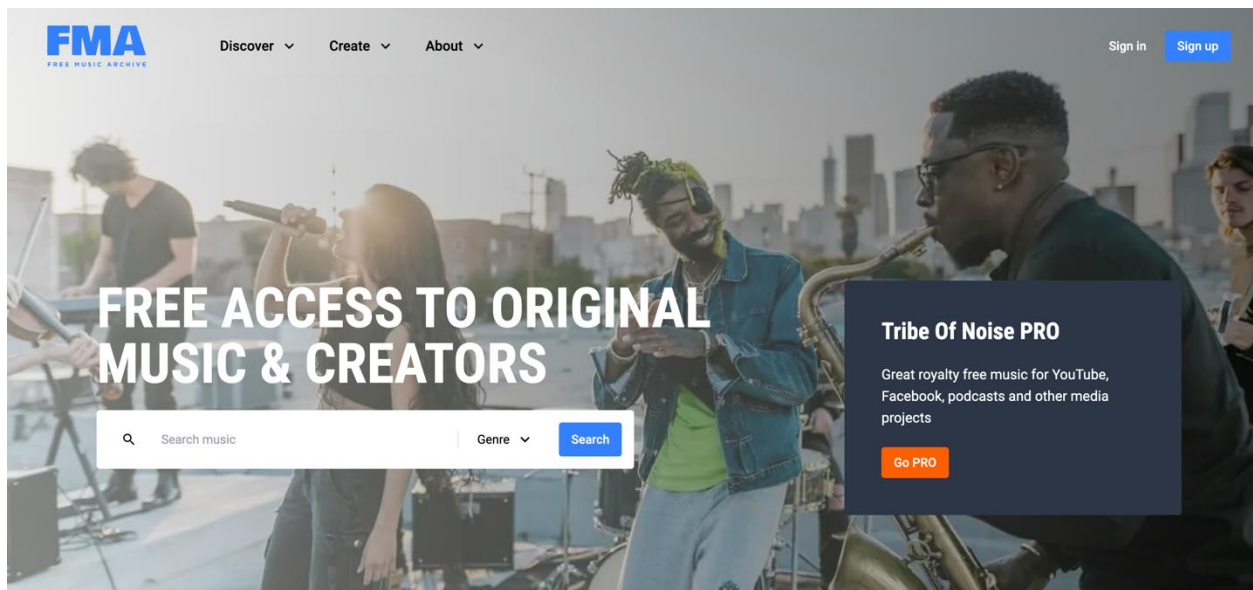


HW Assignment 4 (A4) Part 2: Dimension Reduction of FMA

CS6140: Machine Learning Fall 2024

Due Date: Thursday, _____

(10 points)



Scenario

This assignment is a continuation of the fourth assignment on Unsupervised Learning. This part of the assignment deals with the concept of dimension reduction. Dimension reduction is a technique to reduce the number of features or variables in a dataset while preserving its essential structure and relationships. It aims to simplify complex, high-dimensional data, making it easier to visualize, analyze, and process. Common dimension reduction techniques include Principal Component Analysis (PCA), t-Distributed Stochastic Neighbor Embedding (t-SNE), Uniform Manifold Approximation Projection (UMAP) and others. These methods (except PCA) work by projecting the high-dimensional data onto a lower-dimensional space, retaining as much relevant information as possible while removing noise and redundancy. Dimension reduction can improve computational efficiency and reduce overfitting.

Tasks

This assignment aims to explore and apply various dimension reduction techniques to the Free Music Archive (FMA) dataset. You will use different algorithms to reduce the dataset's dimensionality and visualize the results, comparing the effectiveness of each method. Additionally, you will examine the relationships between the reduced dimensions using correlation analysis.

1. **Data Preparation:** Start with feature extraction on the FMA dataset, ensuring the data is clean and ready for analysis. Standardize or normalize the data as necessary to prepare for dimension reduction.

2. **Dimension Reduction Techniques:** Apply the following dimension reduction techniques to the following domains separately: chroma, mfcc, spectral, tonnetz,

- Principal Component Analysis (PCA)
- Uniform Manifold Approximation and Projection (UMAP)
- t-Distributed Stochastic Neighbor Embedding (t-SNE) **This is optional**
- Locally Linear Embedding (LLE)
- ISOMAP

Use each technique to reduce the dataset to various target dimensions (e.g., 2 or 3 dimensions, except PCA).

3. **Visualization:** For each technique, create visualizations of the dataset in 2D and 3D space. Use color-coding or labeling to distinguish any available categories, such as genres, within the dataset. Compare the visualizations to assess how each method captures clusters or patterns in the data.

4. **Correlation Analysis:** For each dimension reduction technique, calculate and interpret correlations among the reduced dimensions. Identify any strong correlations that might suggest redundancies or patterns in the reduced data.

5. **Comparative Analysis:** Compare the effectiveness of each dimension reduction method in preserving the structure and patterns of the dataset. Discuss which methods capture the most meaningful data representations for the FMA dataset and why.