Vrije Universiteit Amsterdam

Multivariate statistics

Academic year 2024-2025

Individual Assignment 2-C Create a "Song Radio" playlist

Authors:

Duy Phan - (2802376) - (d.j.d.phan@student.vu.nl)

March 20, 2025

1 Introduction

This report focuses on creating a song radio playlist including 20 songs with similar features. This investigation uses the "songlist" dataset that contains 3090 songs to be analyzed using four main characteristics involving danceability, energy, loudness, and tempo. The main task is to assign each song to each cluster through similar features to construct the playlist. This is conducted by using the Expectation Maximization or EM Method to estimate the Gaussian Mixture Model (GMM) with chosen 5 components.

1.1 Estimation of Gaussian mixture model by EM method

A Gaussian mixture model is a soft clustering method where a given data point can belong to a cluster with a particular probability. The Expectation Maximization (EM) algorithm is used to estimate a Gaussian mixture model which is an iterative approach optimizing the Gaussian parameters and the EM algorithm focuses on the following two main steps:

In the Expectation step or E-step, the algorithm computes the posterior probability that each given data point belongs to a cluster depending on the existing estimated parameters, including the mean, covariance matrix, and the mixture weight (μ, Σ, w). This can be done by computing the conditional expectation of the loglikelihood with respect to the unobserved variable Z given the initialized parameters $\theta^{(0)}$.

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{i=1}^{n} \log\left(f_{X,Z}(X_i, Z_i; \theta)\right) \middle| X_i = x_i, \theta^{(0)}\right] \\
&= \sum_{i=1}^{n} \mathbb{E}\left[\sum_{\ell=1}^{L} Z_{i,\ell}\left(\log\left(f_\ell(X_i; \theta_\ell)\right) + \log(w_\ell)\right) \middle| X_i = x_i, \theta^{(0)}\right] \\
&= \sum_{i=1}^{n}\sum_{\ell=1}^{L} \mathbb{E}\left[Z_{i,\ell} \middle| X_i = x_i, \theta^{(0)}\right]\left(\log f_\ell(x_i; \theta_\ell) + \log w_\ell\right) \\
&\quad \text{where} \quad \mathbb{E}\left[Z_{i,\ell} \middle| X_i = x_i, \theta^{(0)}\right] = \mathbb{P}[Z_{i,\ell} = 1 | X_i = x_i, \theta^{(0)}] = \pi_{i,\ell}^{(0)} \\
&= \sum_{i=1}^{n}\sum_{\ell=1}^{L} \pi_{i,\ell}^{(0)}\left(\log f_\ell(x_i; \theta_\ell) + \log w_\ell\right)
\end{aligned}
$$

where

- $x \in \mathbb{R}^n$ is the n dimensional observed random variable
- $\theta^{(0)}$ is a set of initial parameters
- $\mu \in \mathbb{R}^n$ is the mean of k-th Gaussian component.
- $\Sigma \in \mathbb{R}^{n \times n}$ is the covariance matrix of the i-th Gaussian component.
- $w_\ell \in \mathbb{R}$ and $\sum_{\ell=1}^{L} w_\ell = 1$, where w_ℓ is the mixture weights
- $L = 5$ components

This yields the following posterior probability

$$\pi_{i,\ell}^{(0)} = \frac{w_\ell^{(0)} f_\ell(x_i; \theta_\ell^{(0)})}{\sum_{m=1}^{L} w_m^{(0)} f_m(x_i; \theta_m^{(0)})}$$

In the Maximization Step (M-Step), the EM method will update the μ, Σ, w parameters through the computed probabilities in the E-step. This can be conducted by maximizing the expected loglikelihood with respect to θ.

$$\hat{\theta} = \arg\max_{\theta} \left\{ \sum_{i=1}^{n} \sum_{\ell=1}^{L} \underbrace{\pi_{i,\ell}^{(0)}}_{\text{fixed! (i.e. a number)}} (\log f_\ell(x_i; \theta_\ell) + \log w_\ell) \right\} \tag{1}$$

The above optimization problem can be solved by taking a derivative with respect to $\theta = (\mu, \Sigma, w)$ that yields the following updated optimal parameters.

$$\begin{aligned}
\hat{w_\ell} &= \frac{1}{n} \sum_{i=1}^{n} \pi_{i,\ell}^{(0)} \\
\hat{\mu_\ell} &= \frac{\sum_{i=1}^{n} \pi_{i,\ell}^{(0)} x_i}{\sum_{i=1}^{n} \pi_{i,\ell}^{(0)}} \\
\hat{\Sigma_\ell} &= \frac{\sum_{i=1}^{n} \pi_{i,\ell}^{(0)} (x_i - \mu_\ell)(x_i - \mu_\ell)^\top}{\sum_{i=1}^{n} \pi_{i,\ell}^{(0)}}
\end{aligned}$$

Finally, EM repeats to evaluate the E and M-steps and update until the model converges.

1.2 Interpretation of the Estimates of Mixture Components

The estimated means and mixture weights with a chosen L = 5 components are summarized in the following Table 1.

Table 1: Estimated Means and Mixture Weights for Gaussian Mixture Components

Component	Danceability	Energy	Loudness	Tempo	Weight
1	0.12	-0.88	-0.48	-1.08	0.156
2	0.36	-0.22	0.15	0.40	0.309
3	-0.50	0.97	0.55	0.39	0.231
4	-0.30	-1.01	-1.50	0.21	0.131
5	0.14	0.65	0.56	-0.41	0.174

Notes: Since the data is standardized with a mean of 0 and a standard deviation of 1, the average threshold is the standardized mean with a value of 0. Values are above the average if the feature values > 0 and below the average are the values < 0. The substantial deviations are the values around ± 1.

Discussion:

Table 1 shows that for the first component with a weight of 0.156 shows that 15.6 % of the songs have a slight danceability value above the average (0.12) and below-average values of energy and loudness at (-0.88) and (-0.48) respectively, especially the notable low value for tempo (-1.08) and these suggesting songs are less energetic, slow speed and mellow sounds with moderate danceability. These songs are mostly pop songs with a chill melody. Secondly, the second component with 30.9% of weight that accounts for the highest proportion of the dataset that includes songs with surpassing average rates for danceability (0.36) and tempo (0.40) with a slight one that exceeds the mean for loudness at (0.15) while there is negligible falling below average energy values at (-0.22). These values demonstrates that these songs are moderately danceable and fast and slightly loud and low energy, suggesting they are catchy and upbeat tunes such as pop songs. Furthermore, there is 23.1% of songs in the third component which have a relatively loud, extremely energetic and rapid melody showing intense, aggressive and strong musics which are rock and metal due to high energetic value (0.97) which is nearly one and values of loudness and tempo that are larger than standard while the danceability at (-0.50) that is less than the mean. Additionally, component 4 involves songs with soft and gentle tunes, low energy with a moderately fast speed but not very danceable demonstrating soft pop and metal songs because of considerably low values for energy and loudness at (-1.01) and (-1.50) respectively, and there

are values nearly around the standardized midpoint for danceability (-0.30) and tempo (-1.50) and they are accounted for about 13.1% of the datasets which is the lowest proportion. Finally, approximately 17.4% of songs in the fifth component in Table 1 is a bit above the average threshold for danceability at (0.14) but there are moderately higher than the typical values for energy and loudness and tempo is slightly below the mean of zero. This shows the corresponding genres of pop songs, rock and metal ballads that are relatively energetic, and loud but with slow and emotional melodies and somewhat danceability.

Conclusion:

The second component has the largest weight of 30.9% illustrating most of the dataset containing pop songs with catchy and upbeat melodies, as opposed to 13.1% of the fourth component with a calmer tune and less energetic songs. Especially, component 3 which is the second largest proportion of the datasets including rock and metal songs which are highly energetic, loud and aggressive tunes.

1.3 Interpretation of the Clusters between Energy and Loudness

The scatter plot of Figure 1 shows the result of clustering using a Gaussian Mixture Model (GMM). The data points visualize how each song is assigned to one of the five clusters based on the song characteristics including energy and loudness. The data points are divided into five cluster with different colours of purple, blue, yellow, green, and teal, indicating GMM has identified five distinct groups within the dataset.

a) Cluster Separation by Energy and Loudness

Figure 1 illustrates that there is a huge overlap between clusters 1, 2, 3 and 4 with corresponding colors of purple, blue, teal and green in the middle levels of energy and loudness axes where the energy ranging from -1 to 1 and the loudness ranging from -2 to 0 and this makes the grouping separation become less obvious along these axes. Another notable point for this is that the green cluster has the widest spread along axes of energy from -3 to 1 and axes of loudness from -6 to around 0. However, cluster 5 in yellow is located distinctly from other clusters at high energy in between 0 and 1 and loudness levels between -0.5 and 2. Therefore, the five groups are not clearly separated when using measurements of energy and loudness due to the

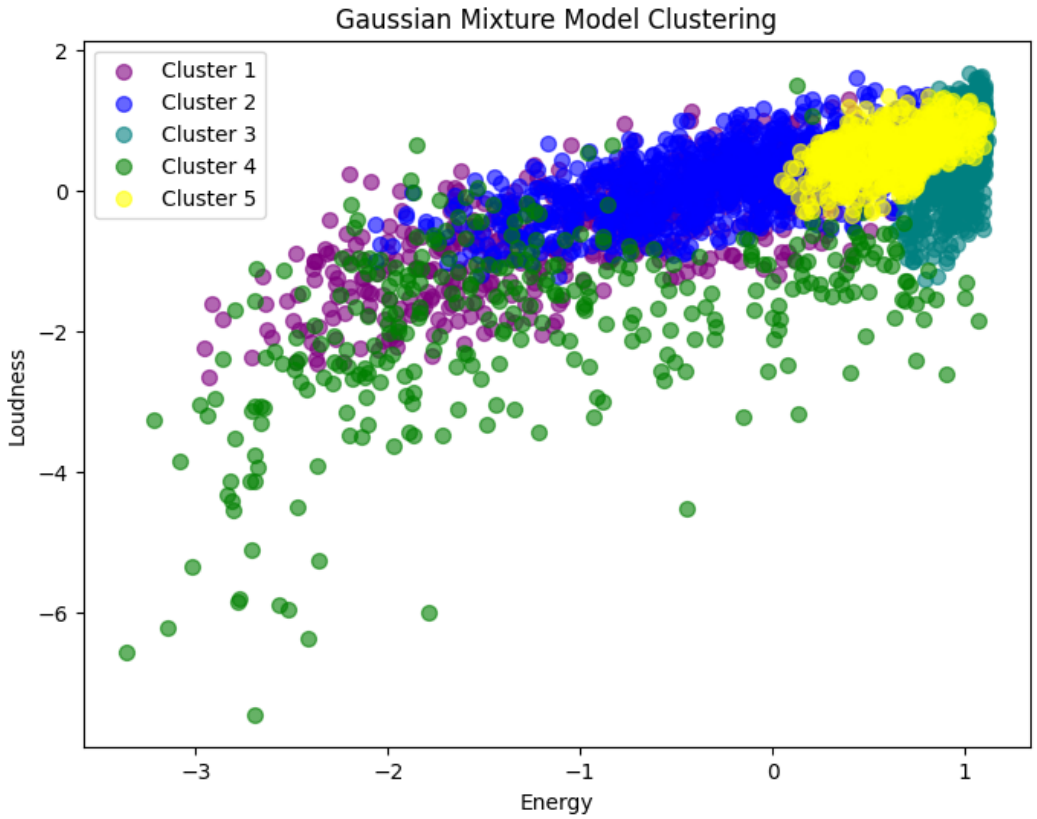

Figure 1: Scatter-plot of Energy vs Loudness
Notes: Purple: Cluster 1, Blue: Cluster 2, Teal: Cluster 3, Green: Cluster 4, Yellow: Cluster 5

overlapping between clusters 1, 2, 3 and 4, especially the substantial expansion of cluster 4 along both axes. The overlapping clusters are highly likely due to the multidimensional space plotting into 2 dimensions.

b) Interpretation of the clusters

For the first cluster in purple represents the softer pop and rock songs since it is centered at the levels where the energy ranges from -2 to 0.5 and the loudness ranges from -2 to 0. The second blue cluster is positioned higher compared to the purple cluster that shows slightly more energetic and louder songs such as rock songs and upbeat pop songs. Furthermore, there is a high-energy song with different levels of loudness in the third cluster this can represent gentle to loud rock and metal songs. The fourth green cluster whose largest spread from low to mid loudness and low to high energy levels indicates soft and calm melodies for pop songs and less intensive rock and metal songs. Finally, the yellow cluster or cluster 5 located separately at considerably high energy and loudness levels suggests the metal and strong rock songs. These can be observed from

Figure 1.

1.4 A Song Radio Playlist

a) Steps to Construct Radio Playlist

The following steps are conducted to create a song radio playlist. Firstly, "Detonation" by Trivium is chosen as a reference song with key characteristics of a metal song to construct the playlist. Then the selected song is checked whether it is found or not, if not, the alternative song will be replaced before processing to the main steps. Secondly, the E-step is applied to calculate the posterior probability which is the probability that each song belongs to each cluster. The GMM is then used to determine which cluster the chosen song belongs to and this is done by identifying the cluster with the highest posterior probability here it is cluster 5 with a posterior probability of nearly 0.604. Next, including only songs that belong to the same cluster the reference song is assigned to. Finally, The posterior probabilities of these songs are sorted in descending order to obtain the final playlist of 20 songs. This song radio playlist is shown in the following Table 2.

b) Evaluation of Result

After listening to several songs such as "Reclamation" by Lamb of God, "Hells Bells" by AC/DC, "This Side of Fate" by Alter Bride, "Forgotten Faces" by Avenged Sevenfold and "I Die in Degrees" by All That Remains. These songs have the same characteristics as the reference song Detonation by Trivium which is a metal song with the same loud, powerful tune, full of energy, with long guitar riffs and intense and strong voices of singing such as screaming. Therefore, the selected songs generally match the style and feel of the selected song, especially in terms of energy and loudness.

Table 2: Song Radio Playlist

Number	Name	Artist	Posterior Probability
1	My Life for Yours	Killswitch Engage	0.960422
2	The Cross	Within Temptation	0.955692
3	Reclamation	Lamb of God	0.952612
4	Still Of The Night	Whitesnake	0.949740
5	Hells Bells	AC/DC	0.949149
6	Forgotten Faces	Avenged Sevenfold	0.948252
7	Sunshine	Aerosmith	0.947153
8	I Die In Degrees	All That Remains	0.946102
9	This Side Of Fate	Alter Bridge	0.940126
10	In the Middle of the Night	Within Temptation	0.939232
11	Let Us Burn	Within Temptation	0.937776
12	Stars Dance	Selena Gomez	0.936896
13	Misunderstood	Bon Jovi	0.936184
14	We All Fall Down	Bon Jovi	0.935049
15	Endless Nights	Sabaton	0.934451
16	Shot in the Dark	Within Temptation	0.934214
17	This Probably Won't End Well	All That Remains	0.928633
18	The Red	Disturbed	0.924472
19	Supernova	Within Temptation	0.924398
20	Panzerkampf	Sabaton	0.924005