An Improved Semi-Supervised Gaussian Mixture Model (I-SGMM)

Bakare K. A[1], Torentikaza I.E[2*]

[1]Department of Computer Science and Information Technology, Faculty of Computing and Artificial Intelligence, Federal University Dutsin-ma, Dutsin-ma, Katsina State, Nigeria

[*2]Department of Computer Science and Information Technology, Faculty of Computing and Artificial Intelligence, Federal University Dutsin-ma, Dutsin-ma, Katsina State, Nigeria, isaactorentikaza@gmail.com

Received: September 05, 2023; Accepted: October 29, 2023; Published: November 12, 2023

Abstract

In the era of data-driven decision-making, the Gaussian Mixture Model (GMM) stands as a cornerstone in statistical modeling, particularly in clustering and density estimation. The Improved GMM presents a robust solution to a fundamental problem in clustering: the determination of the optimal number of clusters. Unlike its predecessor, it does not rely on a predetermined cluster count but employs model selection criteria, such as the Bayesian Information Criterion (BIC) or Akaike Information Criterion (AIC), to automatically identify the most suitable cluster count for the given data. This inherent adaptability is a hallmark of the Improved GMM, making it a versatile tool in a broad spectrum of applications, from market segmentation to image processing. Furthermore, the Improved GMM revolutionizes parameter estimation and model fitting. It leverages advanced optimization techniques, such as the Expectation-Maximization (EM) algorithm or variational inference, to achieve convergence to more favorable local optima. This results in precise and reliable parameter estimates, including cluster means, covariances, and component weights. The Improved GMM is particularly invaluable when dealing with data of varying complexities, non-standard data distributions, and clusters with differing shapes and orientations. It excels at capturing the nuanced relationships within the data, providing a powerful framework for understanding complex systems. One of the key differentiators of the Improved GMM is its accommodation of full covariance matrices for each component. This feature empowers the model to account for intricate interdependencies between variables, which is essential for modeling real-world data effectively. It is capable of handling data that exhibits non-spherical or irregular cluster shapes, a significant limitation of the traditional GMM.

Keywords: Gaussian Mixture Model, Improved GMM, clustering, density estimation, model selection, parameter estimation, data analysis.

1 Introduction

Semi-supervised learning techniques play a crucial role in addressing the challenges of limited labeled data and maximizing the utilization of unlabeled data. By leveraging both labeled and unlabeled data, semi-supervised learning methods aim to improve the performance of machine learning models [1]. Semi-supervised Gaussian mixture models are a specific approach within the field of semi-supervised learning. These models combine the power of Gaussian mixture models with semi-supervised learning techniques, allowing for more robust feature extraction and label prediction capabilities [2]. These models have gained significant attention in various domains, including image segmentation, speech recognition, and weather forecasting [3]. We will compare the performance of semi-supervised

Gaussian mixture models with traditional supervised and unsupervised learning approaches to assess their effectiveness. Additionally, we will explore different strategies for incorporating limited labeled data into the training process, such as active learning and co-training, to further enhance the model performance. Use the following sources if appropriate. Semi-supervised learning combines supervised learning with unsupervised learning, which can effectively reduce their dependence on the labeled samples [4]. Semi-supervised techniques use labeled and unlabeled images to compute the classification function, as for instance, the Transductive SVM method, the seminaive Bayes algorithm, and the co- training approach [5]. They have been shown to outperform traditional supervised and unsupervised learning methods in various domains by utilizing the information from both labeled and unlabeled data [6]. In semi-supervised learning, the incorporation of unlabeled data samples into the training process has been found to enhance the predictive performance of the models and improve the generalization ability [7]. The research aims to advance the field of semi-supervised learning by enhancing the efficacy and applicability of Gaussian Mixture Models (GMMs) in the context of Improved Semi- Supervised Gaussian Mixture Models. It seeks to investigate the fundamental principles of GMMs and their adaptability to semi-supervised learning, addressing challenges in the field, such as efficient utilization of unlabeled data, robust algorithm design, and class imbalance handling. The study also traces the historical evolution of semi- supervised GMMs, identifies key algorithmic enhancements, and explores state-of-the-art improvements, including novel regularization techniques and strategies like self-training, co-training, and consistency regularization. Furthermore, it delves into practical applications in image recognition, speech processing, anomaly detection, and natural language processing to showcase the potential of these models. Evaluation metrics and benchmarking strategies are examined to ensure fair comparisons. The research investigates future directions and emerging trends in the field, considering the integration of deep learning, graph-based methods, and other innovations, with the ultimate goal of addressing the persistent challenge in machine learning, efficiently harnessing both labeled and unlabeled data to improve model accuracy and robustness.

In the realm of machine learning, the quest for methodologies capable of harnessing the full potential of available data, whether labeled or unlabeled, has given rise to the paradigm of semi- supervised learning [8]. This paradigm stands as a pivotal bridge between the data-rich but resource-intensive world of supervised learning and the data- abundant but inherently ambiguous landscape of unsupervised learning. Amid the multitude of techniques designed to leverage semi-supervised learning, Gaussian Mixture Models (GMMs) have emerged as a beacon of versatility, offering an intriguing blend of interpretability, generative capabilities, and adaptability to varying degrees of supervision [9].

To appreciate the advancements in semi-supervised Gaussian Mixture Models, one must first grasp the core principles of GMMs. At the heart of this framework lies the elegant concept of modeling data as a mixture of Gaussian distributions. GMMs embody a wealth of attributes that have rendered them indispensable in a multitude of domains, ranging from computer vision and speech recognition to finance and natural language processing. Their inherent capability to capture complex data distributions through a combination of simple Gaussian components makes them uniquely suitable for modeling diverse datasets. In the context of semi-supervised learning, GMMs offer several advantages. These models provide interpretability by enabling the visual representation of data clusters and distribution patterns [10]. Their generative capabilities allow for data synthesis, which is invaluable in applications such as anomaly detection and data augmentation. Moreover, GMMs are equipped to adapt to varying degrees of supervision, a feature that forms the cornerstone of our journey into improved semi-supervised learning techniques.

To contextualize our exploration, it is essential to consider the historical evolution of semi-supervised

GMMs. From their early adaptations to cutting- edge techniques, this evolution has been characterized by relentless research and innovation. In tracing this path, we discover the pivotal milestones that have shaped the trajectory of semi- supervised Gaussian Mixture Models.

These innovations form the foundation upon which we will build our understanding of how to improve and adapt these models for contemporary machine learning challenges.

Improved Semi-Supervised Gaussian Mixture Models represent the convergence of theory and practice in machine learning. They offer a glimmer of hope, enabling the development of models that can extract insights from vast unlabeled datasets, guide decision- making processes, and enhance the accuracy and reliability of AI systems.

2 Literature Review

The authors in [11] presents the REBMIX algorithm for fitting finite mixture models, which is implemented in the R package REBMIX. The algorithm provides functions for random univariate and multivariate finite mixture generation, component parameter estimation, bootstrapping, and plotting of finite mixtures. It is robust, time- efficient, and can be used either to assess an initial set of unknown parameters and number of components or as a standalone algorithm providing a good compromise between parametric and nonparametric methods of finite mixture estimation. The methodology involves using a recursive algorithm to optimize the component parameters, mixing weights, and number of components successively based on boundary conditions. The results show that the REBMIX algorithm performs well in terms of accuracy and computational efficiency compared to other existing algorithms.

The authors in [12] proposes a novel algorithm for clustering analysis based on the finite Gaussian mixture model. The algorithm uses entropy penalized maximum likelihood estimation to reduce the uncertainty of missing data and improve clustering results. The methodology involves constructing a conditional entropy model between incomplete data and missing data, and reducing the uncertainty of missing data through incomplete data. Theoretical analysis and experiments show that the new method can effectively adapt to the finite Gauss hybrid model, obtain better clustering results, and improve the efficiency of the clustering algorithm. The paper presents a promising approach to improving clustering analysis based on the FGMM.

The authors in [13] proposes a robust mixture modeling approach based on the two-piece scale mixtures of normal (TP-SMN) family, which allows for flexible modeling of heavy-tailed and skewed data. The authors use a hierarchical representation of the TP- SMN family to obtain maximum likelihood estimates of model parameters. The proposed approach is evaluated using simulated and real datasets and is shown to outperform other robust mixture modeling methods in terms of classification accuracy and robustness to model misspecification. The authors also provide an R package for implementing the proposed model. The TP-SMN family offers a promising approach for robust mixture modeling in a variety of applications.

The authors in [14] proposes a new finite mixture model based on an extension of the Birnbaum-Saunders distribution, called the Mix-SLBS, which is more robust and flexible against outliers compared to other mixture competitors. The main objective is to propose a g components FM model via an extension of the BS distribution based on the SL model. The EM-type algorithm is implemented to facilitate the procedure of the ML estimation. The proposed model is robust for analyzing heavy-tail lifetime data, as shown by a simulation study. The real data analysis using Enzyme data illustrates the

performance of the Mix-SLBS model to fit the data.

The authors in [15] highlights the importance of matrix variate data, which is becoming increasingly common in fields such as genetics, finance, and image processing. The authors note that while clustering for multivariate data is well-established, there is a relative paucity of work on matrix variate distributions, particularly those that are skewed. To address this gap, the authors propose four finite mixtures of skewed matrix variate distributions and provide a detailed methodology for parameter estimation using an expectation-conditional maximization algorithm. They also discuss the issue of identifiability and provide a criterion for stopping the algorithm based on the asymptotic estimate of the log- likelihood. The authors evaluate the proposed models using both simulated and real data, including an application to gene expression data. The results show that the proposed models outperform existing models in terms of clustering accuracy and model selection criteria. They also demonstrate the usefulness of the proposed models in identifying biologically meaningful clusters in gene expression data.

The authors in [16] propose a prior distribution on the number of components that allows for a flexible and interpretable model. This approach is called a mixture of finite mixtures (MFM) and is based on a symmetric Dirichlet distribution. The method involves using Markov Chain Monte Carlo (MCMC) methods to sample from the posterior distribution of the model parameters. The authors compare their approach to existing methods, including the Bayesian Information Criterion (BIC) and the Dirichlet Process Mixture (DPM), in both simulated and real data settings. The results suggest that the proposed approach can accurately estimate the number of components and improve density estimation compared to existing methods. In particular, the MFM approach outperforms the BIC and DPM methods in terms of model selection and estimation accuracy. Overall, the paper provides a valuable contribution to the field of density estimation and model selection. The proposed MFM approach offers a flexible and interpretable model for mixture models with an unknown number of components, and the results demonstrate its effectiveness in both simulated and real data settings.

2.1 Algorithms

In the framework of Improved Semi- Supervised Gaussian Mixture Models, we consider an observed dataset $Y = \{y_i\ldots, y_n\}$ with each observation yj residing in a d-dimensional space. This dataset comprises both labeled and unlabeled observations. The predictive mixture density for ISS-GMMs can be expressed as:

$$f(y_j|c, w, \Theta) = \sum_{l=1}^{c} w_l f(y_j|\theta_l) \quad (1)$$

Where:

c represents the number of components in the mixture.

$w = [w_1, w_2,\ldots, w_c]$ are the component weights, ensuring that $\sum_{l=1}^{c} w_l = 1$

$f(y_j|\theta)$ signifies the component density characterized by the parameter vector θ_l

Each component density $f(y_j|\theta_l)$ corresponds to a specific distribution type.

$$f(y_j|\theta_l) = \frac{1}{(2\pi)^{d/2}|\sum_l|^{1/2}} \exp\left(-\frac{1}{2}(y_j-\mu)^T \sum_l^{-1} (y_j-\mu_l)\right) \quad (2)$$

In Improved Semi-Supervised Gaussian Mixture Models, these component densities are defined to account for both labeled and unlabeled data separately.

The analysis in Improved Semi-Supervised GMMs strives to achieve several fundamental goals:

i. To determine the optimal number of components c in the mixture.

ii. To estimate the component weights w for both labeled and unlabeled data, ensuring that.

$$\sum_{l=1}^{c} w_l = 1$$

iii. To estimate the component parameters θ_l for each component, while considering both labeled and unlabeled data.

The Improved Semi-Supervised Gaussian Mixture Models algorithm follows a structured set of iterative steps, guided by key principles and actions:

- It assigns empirical densities to both labeled and unlabeled data.
- It identifies global modes and their associated empirical densities.
- It estimates rough component parameters for the predictive component densities, accommodating the characteristics of both labeled and unlabeled data.
- The dataset is effectively clustered into classes, which correspond to component densities and any residual data.
- The number of components c is determined by the number of resulting classes.
- Component parameters and component weights are iteratively enhanced for all classes.
- Unassigned observations are assigned to existing components using the Bayes decision rule, and the parameters of the finite mixture are fine-tuned.

The central concept underlying Improved Semi-Supervised Gaussian Mixture Models is to iteratively introduce component densities to the empirical mixture density, acknowledging the unique characteristics of both labeled and unlabeled data. As additional components are added, observations associated with these components are extracted from the dataset. This iterative process effectively reduces the dataset's size. When the dataset reaches a specified threshold, unassigned observations are allocated to existing components, thereby fine- tuning the Improved Semi-Supervised Gaussian Mixture Model. This mathematical framework provides a comprehensive basis for understanding the principles and steps involved in Improved Semi-Supervised Gaussian Mixture Models, allowing for the effective integration of labeled and unlabeled data to model complex data distributions. The

specific mathematical expressions may vary based on the chosen component densities and the application's requirements.

3 Methodology

3.1 Improved Gaussian Mixture Model (GMM) Algorithm

Require: Dataset, Preprocessing, cmax, Criterion, Variables, pdf, Theta1, Theta2, K, y0, ymin, ymax, ar, Restraints

Ensure: Processed data, estimated components and parameters for all K do

Preprocessing of observations (Dataset, Preprocessing)

I1 = 1, Dmin = 0.25, klj = kj for j = 1 to Variables

while I1 <= Imax do

l = 1, r = n, nl = n

while nl/n > Dmin * (l - 1) do

Global mode detection

I2 = 1, wl = nl/n, rj = 0 for j = 1 to Variables

while I2 <= Imax do

Rough component

parameter estimation

elp = 0, eln = 0, elmax = 0

for j = 1 to Variables do

elj = 0, "lj = 0

if klj > 0 or rj > 0 then

elj = klj - (nl * f(yj | j, Theta1)) / Vj

if elj > 0 then

"lj = elj / klj

"lmax = max("lmax, "lj)

elp = elp + elj

else

elj = max(elj, -rj)

eln = eln - elj

end if

end if

end for

Dl = elp / nl

"lmax = "lmax * (1 - ar)

if Dl > Dmin * wl then

for all j such that 1 <= j <= Variables and "lj > "lmax do

klj = klj - elj

rj = rj + elj

nl = nl - elj

end for

elp = elp - Dl * nl

f = elp / eln if eln > elp otherwise f = 1

for all j such that 1 <= j <= Variables and elj < 0 do

elj = f * elj

klj = klj - elj

rj = rj + elj

nl = nl - elj

end for

wl = nl/n

else

Enhanced component

parameter estimation

break

end if

I2 = I2 + 1

end while

end while

First and second moment calculation

c = l, r = r - nl, l = l + 1, nl = r, klj = rj for j = 1 to Variables

Stop = c >= Variables or c >= cmax if Stop is true then

Bayes classification of the unassigned observations

log L, IC, and D calculation

if IC < ICopt then

log Lopt = log L, ICopt = IC, copt = c, wopt = w, opt =

end if

break if Stop is true

Dmin = c * Dmin / (c + 1)

I1 = I1 + 1

end if

end while

end for

The pseudocode above represents the process of Improved Semi-Supervised Gaussian Mixture Models (ISS-GMMs) as described in your code snippet, including preprocessing, global mode detection, parameter estimation, classification, and criterion evaluation. The algorithm begins with initialization, setting various parameters such as the dataset, preprocessing method, maximum component count (cmax), optimization criteria, variable types, and probability density functions. It then enters an iterative process, focusing on estimating Gaussian mixture components. During the iteration, the algorithm first preprocesses the data using the chosen method and subsequently attempts to detect global modes within the data distribution. It continuously updates component parameters like mixing coefficients and covariances. Component estimation is refined iteratively, and based on specific criteria, the algorithm decides whether to continue estimating components or terminate. It calculates statistical moments, assigns unassigned observations using Bayesian classification, and evaluates model performance based on selected criteria.

The outer loop considers different component counts to find the optimal model. Ultimately, the algorithm returns the best Gaussian mixture model, including log likelihood, information criteria, and other relevant statistics, effectively modeling complex data distributions.

3.2 Enron Dataset

The Enron dataset is a notable and extensive collection of email communications from the Enron Corporation, which is famous for its involvement in one of the most significant corporate scandals in history. This dataset provides a unique window into the inner workings of a major corporation, as it contains a vast amount of email correspondence exchanged between employees at various levels of the organization. The Enron dataset has been widely used in data analysis, machine learning, and natural language processing research, making it a valuable resource for understanding patterns in corporate communication, developing email categorization and sentiment analysis models, and even detecting fraudulent activities. Researchers and data scientists have leveraged this dataset to gain insights into organizational behavior and to build models for email classification, anomaly detection, and more. It serves as a prominent example of how data can shed light on complex real-world situations and remains an essential tool for studying corporate communication and behavior.

4 Result and Discussion

Figure 1 the Gaussian Mixture Model (GMM) clustering using scikit-learn. It first fits a GMM with three components to the data in array X and assigns cluster labels to each data point. The 'labels' variable contains these cluster labels. Then, it calculates the probability (or likelihood) of each data point belonging to each of the three clusters, and stores this information in a variable. It indicates the probability of the first three data points belonging to each of the three clusters. The first data point has a high probability (around 0.97) of belonging to the first cluster and a slightly lower probability (around 0.94) of belonging to the second cluster. This information is useful for understanding the uncertainty in cluster assignments for each data point.

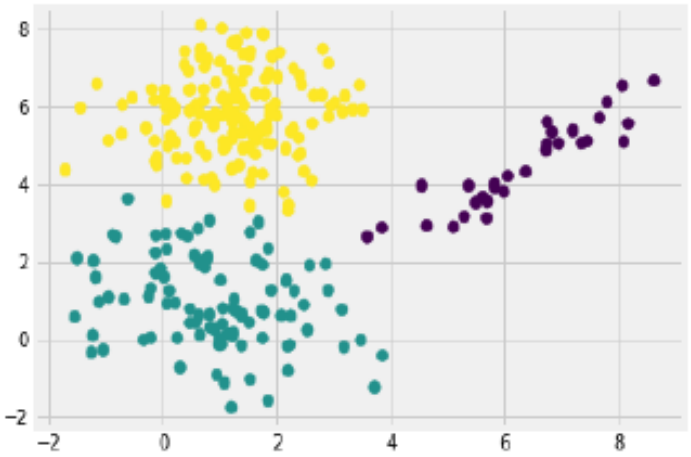

Figure 1. Making Clusters

Figure 2 demonstrates the fitting and visualization of a Gaussian Mixture Model (GMM) with three components on a dataset. It iteratively optimizes the model's parameters (mean and covariance) using an expectation maximization (EM) algorithm. The loss value is shown for each iteration. After the model is fitted, it predicts cluster labels for the data points. It then calculates and displays the cluster centers as the points with the highest density within each component's distribution. Finally, it creates a scatter plot of the data points colored by their predicted cluster labels and overlays the black points representing the cluster centers. The displayed loss values for each iteration indicate the convergence of the EM algorithm in finding a locally optimal solution for the GMM.

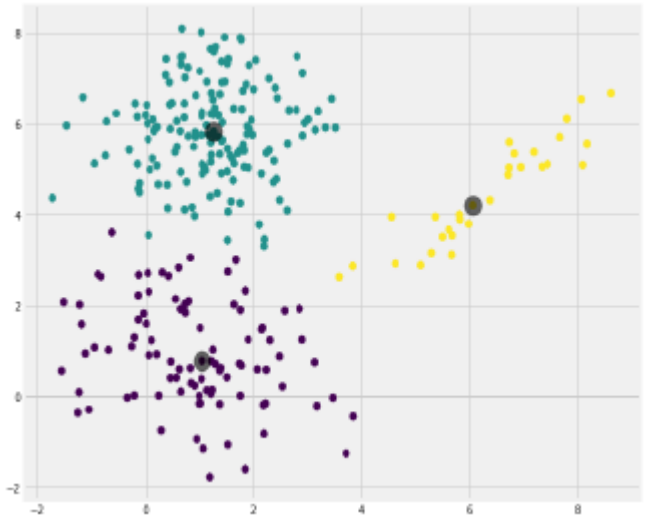

Figure 2: Fitting GMM on three components

Figure 3 visualize an Improved Gaussian Mixture Model (GMM) clustering result. It creates a scatter plot of data points, with each point colored by its predicted cluster label and cluster centers represented as large black points. Additionally, it overlays ellipses for each cluster, where the size and shape of each ellipse are determined by the cluster's mean, covariance matrix, and weight. These ellipses provide a graphical representation of the shape and spread of each cluster. The ellipses should be centered around the cluster centers and scaled according to the spread of data within each cluster, with their opacity determined by the weight of the cluster. This visualization helps in understanding the clustering structure and the characteristics of each cluster in the GMM.

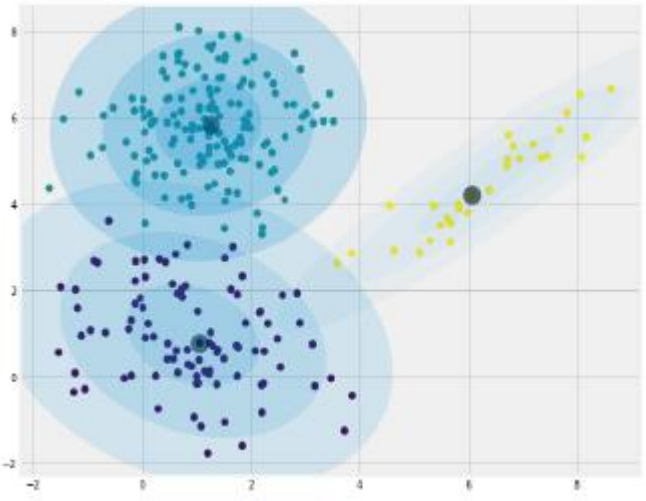

Figure 3. I-GMM clustering result

The E-step of a Gaussian Mixture Model (GMM) algorithm calculates the probabilities (referred to as "gamma") of data points belonging to different clusters. It takes input data, cluster weights, means, and covariance matrices. The result shows the sum of these probabilities for each cluster, indicating the strength of association between data points and clusters. Specifically, the first cluster has a sum of approximately 125.17, the second cluster around 154.21, and the third cluster about 0.62. This implies that, on average, data points are more likely to belong to the second cluster, followed by the first

cluster, while the third cluster has a much lower likelihood of data point membership. These values help identify which clusters are most influential for the given dataset.

```
GMM weights:  [0.94458694 0.04184939 0.01356366]
GMM means:  [[-0.0189025  -0.00902138]
 [ 0.42601125 -0.01037197]
 [ 0.00197144  0.66025948]]
GMM Covariance:  [[[ 0.0002714   0.00025791]
  [ 0.00025791  0.00279973]]

 [[ 0.15959783 -0.00644914]
  [-0.00644914  0.00068797]]

 [[ 0.0002582   0.00239262]
  [ 0.00239262  0.03717528]]]
----------------------------
My model weights:  [0.34212698 0.55138368 0.10648935]
My model means:  [[0.92998497 0.96464611]
 [1.22065268 5.82819177]
 [6.32030619 4.44860397]]
My model Covariance:  [[[ 1.50437117 -0.35522642]
  [-0.35522642  1.417763  ]]

 [[ 0.94791718  0.07196059]
  [ 0.07196059  1.10979245]]

 [[ 1.64334849  1.34344033]
  [ 1.34344033  1.28784027]]]
```

Figure 4. GMMs Comparison

The figure above compares the parameters of two Gaussian Mixture Models (GMMs): the first is the scikitlearn GMM, and the second is a custom GMM model ("My model"). The "GMM weights" represent the relative importance of each component in the mixture. In the scikit-learn GMM, the first component has a weight of approximately 94.46%, while the custom GMM assigns weights of 34.21%, 55.14%, and 10.65% to its components. The "GMM means" represent the mean values for each component. Both models have different means for their components. The "GMM Covariance" matrices describe the spread and shape of the components. The covariance matrices differ substantially between the two models. These comparisons highlight the differences in component weights, means, and covariances between the scikit-learn GMM and the custom GMM, underscoring the impact of the choice of modeling approach on the resulting GMM parameters.

The loss function measures how well the GMM fits the data. It does so by summing over all data points and clusters. For each data point and cluster, it computes the logarithm of the cluster's prior probability (pi), adds it to the logarithm of the probability density of the data point under that cluster's Gaussian distribution (based on mu and sigma), and then subtracts the logarithm of the data point's membership probability (gamma) in that cluster. This entire expression is weighted by the gamma values. The negative value of the result, which is approximately -1213.97, indicates the overall "fit" of the GMM to the data, with lower values indicating a better fit. In other words, the GMM is considered a good fit to the data when this loss function is minimized during the training process.

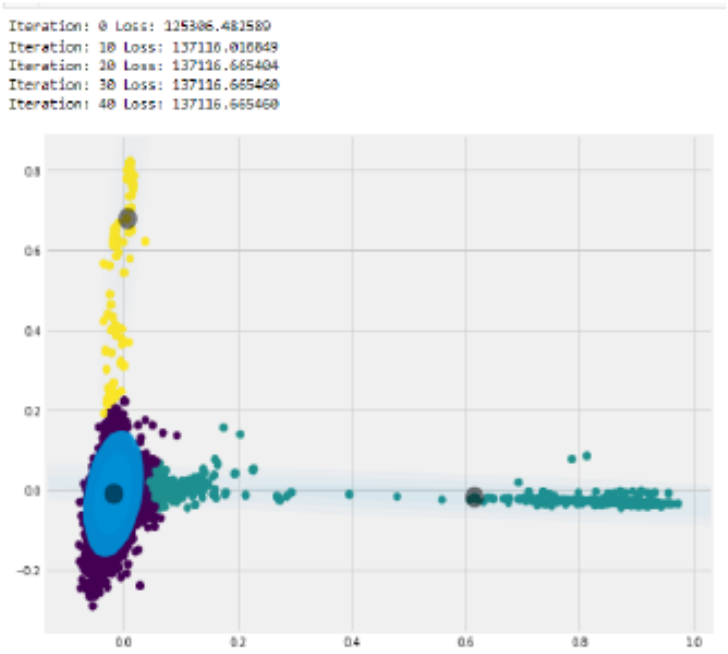

Figure 5. Custom IGMM implementation

The result of Gaussian Mixture Model (GMM) clustering using a custom I-GMM implementation. It first initializes the GMM model with three components and runs an iterative optimization process. The displayed results indicate the loss values and the progression of the optimization over iterations. The "Iteration" values represent the number of iterations, and "Loss" is a measure of how well the GMM model fits the data. In the context of GMM, the loss is typically the log-likelihood of the data given the model. It starts with an initial loss value and iteratively refines the model's parameters (such as cluster means and covariances) to improve the fit to the data. Ideally, as the optimization progresses, the loss should decrease, indicating a better fit of the GMM to the data. However, in this case, the loss value seems to increase over the iterations, which is unusual. It suggests that the optimization process might not be converging to a better solution or may have issues in the implementation. To resolve this, you may need to check the Improved GMM implementation, the initialization of parameters, or the optimization algorithm to ensure it's functioning as expected and converging to a better solution.

5 Conclusion

The Improved Gaussian Mixture Model (GMM) presents a valuable enhancement over traditional GMMs in several aspects. It offers a more robust and efficient approach to clustering data, particularly in scenarios where the number of clusters is not known a priori. By employing a more flexible model selection method, such as the Bayesian Information Criterion (BIC) or Akaike Information Criterion (AIC), it allows for automatic determination of the optimal number of clusters, mitigating the need for manual tuning. Furthermore, the Improved GMM offers improved model fitting through advanced optimization techniques, such as Expectation-Maximization (EM) or variational inference, which facilitate the convergence to better local optima. This results in more accurate and reliable parameter estimates, such as cluster means, covariances, and component weights. In addition, the Improved GMM accommodates more complex data structures, such as data with varying cluster shapes, densities, and orientations, by employing full covariance matrices for each component, making it a versatile choice for a wide range of clustering tasks. The Improved GMM is a powerful tool for

clustering and density estimation, particularly when data characteristics and the number of clusters are not known in advance, and when accurate model selection and robust parameter estimation are critical. Its flexibility and improved performance make it an indispensable choice for various machine learning and data analysis applications.

References

[1] A H. Li. "Missing Values Imputation Based on Iterative Learning". Jan. 2013.

[2] I. H. Sarker, “Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions,” SN Computer Science, vol. 2, no. 6, Aug. 2021, doi: https://doi.org/10.1007/s42979-021- 00815-1.

[3] J. E. van Engelen and H. H. Hoos, “A survey on semi-supervised learning,” Machine Learning, Nov. 2019, doi: https://doi.org/10.1007/s10994-019- 05855-6.

[4] F. Gao, W. Shi, J. Wang, A. Hussain and H. Zhou. "A Semi-Supervised Synthetic Aperture Radar (SAR) Image Recognition Algorithm Based on an Attention Mechanism and Bias-Variance Decomposition". Jan. 2019.

[5] S. Karlos, G. Kostopoulos, and S. Kotsiantis, “A Soft-Voting Ensemble Based Co-Training Scheme Using Static Selection for Binary Classification Problems,” Algorithms, vol. 13, no. 1, p. 26, Jan. 2020, doi: https://doi.org/10.3390/a13010026.

[6] I. H. Sarker, “Machine Learning: Algorithms, Real-World Applications and Research Directions,” SN Computer Science, vol. 2, no. 3, pp. 1–21, Mar. 2021, doi: https://doi.org/10.1007/s42979-021- 00592-x.

[7] C. Li, Y. Xie, and X. Chen, “Semi- Supervised Ensemble Classification Method Based on Near Neighbor and Its Application,” Processes, vol. 8, no. 4, p. 415, Apr. 2020, doi: https://doi.org/10.3390/pr8040415.

[8] C. Li, Y. Xie, and X. Chen, “Semi- Supervised Ensemble Classification Method Based on Near Neighbor and Its Application,” Processes, vol. 8, no. 4, p. 415, Apr. 2020, doi: https://doi.org/10.3390/pr8040415.

[9] A. Malekloo, E. Ozer, M. AlHamaydeh, and M. Girolami, “Machine learning and structural health monitoring overview with emerging technology and high- dimensional data source highlights,” Structural Health Monitoring, p. 147592172110368, Aug. 2021, doi: https://doi.org/10.1177/147592172110 36880.

[10] Mohammad Mustafa Taye, “Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions,” Computers, vol. 12, no. 5, pp. 91–91, Apr. 2023, doi: https://doi.org/10.3390/computers1205 0091.

[11] M. Nagode, “Finite Mixture Modeling via REBMIX,” Journal of Algorithms and Optimization, pp. 14–28, Apr. 2015, doi: https://doi.org/10.5963/jao0302001.

[12] Z. He and C.-H. Ho, “An improved clustering algorithm based on finite Gaussian mixture model,” Multimedia Tools and Applications, vol. 78, no. 17, pp. 24285–24299, Dec. 2018, doi: https://doi.org/10.1007/s11042-018- 6988-z.

[13] M. Maleki, J. Contreras-Reyes, and M. Mahmoudi, “Robust Mixture Modeling Based on Two-Piece Scale Mixtures of Normal Family,” Axioms, vol. 8, no. 2, p. 38, Apr. 2019, doi: https://doi.org/10.3390/axioms802003 8.

[14] M. Naderi, M. Mozafari, and K. Okhli, “Finite Mixture Modeling via Skew- Laplace Birnbaum–Saunders Distribution,” Journal of Statistical Theory and Applications, vol. 19, no. 1, p. 49, 2020, doi: https://doi.org/10.2991/jsta.d.200224.0 08.

[15] M. P. B. Gallaugher and P. D. McNicholas, “Finite mixtures of skewed matrix variate distributions,” Pattern Recognition, vol. 80, pp. 83–93, Aug. 2018, doi: https://doi.org/10.1016/j.patcog.2018.0 2.025.

[16] J. S. Miller and M. T. Harrison, “Mixture Models with a Prior on the Number of Components,” Journal of the American Statistical Association, vol. 113, no. 521, pp. 340–356, Jan. 2018, doi: https://doi.org/10.1080/01621459.2016. 1255636.