Journal of Sciences, Computing and Applied Engineering Research (JSCAER), Vol. 1, No.1, pp. 10-19

Published Online (https://jcaes.net) on May 7, 2025 by SciTech Network Press

Machine Learning Based on Exhaustive GMM Clustering Algorithm for Optimal Learning of 5G-NR SINR Datasets

Emughedi Oghu

PhD student, Department of Computer Science, Lead City University, Ibadan OYO State, Nigeria

Received: 30 Feb 2025; Revised: 04 May 2025; Accepted: 05 May 2025; Published: 07 May 2025

Abstract: The rapid evolution of 5G technology has ushered in a new era of unparalleled connectivity and data speeds, creating vast opportunities for innovation across industries. Central to the optimization of 5G networks is the analysis of (Signal-to-Interference-plus-Noise Ratio) quality datasets), which plays a crucial role in understanding network performance and enhancing user experience. In this context, the application of advanced clustering techniques, such as the Gaussian Mixture Model (GMM), offers a powerful framework for extracting valuable insights from 5G-NR SINR datasets. However, the optimal tuning of GMM parameters poses significant challenges that require innovative solutions. This paper explores an exhaustive GMM tuning algorithm designed to achieve optimal cluster learning of 5G-NR SINR datasets, providing a comprehensive analysis of methodologies, results, and implications for network optimization.

Keywords: Machine learning, GMM, SINR, clutering, Exhaustive Algorithm

1. Introduction

Fifth generation 5G-NR (5th Generation New Radio) quality datasets are essential in analyzing the performance of 5G networks. These datasets provide valuable insights into the general quality experienced by users, helping to optimize network performance and quality of service.
The 5G-NR SINR quality refers to the measurement of the signal strength relative to the interference and noise levels in a 5G-NR network. It serves as a crucial metric in assessing network performance and determining the reliability of wireless communication services [1-4].

The paper delves into the intricacies of cluster learning in the context of 5G-NR Signal-to-Interference-plus-Noise Ratio (SINR) quality datasets. With the advent of 5G technology, understanding and optimizing SINR quality is paramount for ensuring efficient and reliable wireless communication networks. Studying SINR Quality in 5G-NR networks is vital for understanding and improving network efficiency, coverage, and capacity. By analyzing SINR Quality datasets, network operators can optimize resource allocation, enhance user experience, and ensure seamless connectivity in the 5G landscape.

Clustering 5G-NR SINR Quality datasets poses challenges due to the high-dimensional nature of network data, presence of noise, and variability in signal strength. These factors can affect the clustering performance and necessitate careful parameter tuning to overcome clustering complexities [2,5].

Specifically, this paper explores the application of Gaussian Mixture Model (GMM) for cluster learning, emphasizing the significance of exhaustive GMM tuning in analyzing and deriving optimal insights from complex 5G-NR SINR datasets. Through a comprehensive methodology and implementation of the tuning algorithm, this study aims to provide valuable insights for enhancing cluster learning processes in the realm of 5G network performance analysis. By clustering SINR data, network operators can identify clusters of high interference levels and take steps to mitigate them through interference cancellation techniques or frequency re-allocation strategies. This can help improve the overall network capacity and reliability, leading to a better user experience and increased customer satisfaction.

2. Theoretical Framework

In wireless communication networks, particularly the 5G-NR mobile broadband, the SINR is a crucial metric that determines the quality of the received signal [1]. In order to optimize the performance of the network, it is important to accurately estimate the SINR values at different locations. One approach to achieve this is through Gaussian Mixture Model (GMM) cluster learning of SINR data.

GMM clustering based analysis is a unique unsupervised learning technique within the realm of machine learning. This method can be utilized to explore and uncover implicit internal patterns present in raw data, making it particularly well-suited for analyzing SINR data. By automatically dividing a data sample set and its sample points into multiple clusters, the clustering method aids in identifying these patterns.

In recent years, the field of cluster analysis has encountered increasingly complex challenges due to advancements in information science and technology. As data dimensions continue to expand, researchers have sought innovative solutions to enhance the effectiveness of cluster analysis. For instance, in [5], Torre integrated dimensionality reduction techniques with clustering by initially clustering data using K-means, followed by projecting the data to a lower dimension to maximize variance among groups.

Furthermore, in [6] and [7], the authors introduced deep embedding clustering, a novel approach that leverages deep neural networks to conduct nonlinear feature extraction for cluster analysis. This cutting-edge method has shown promise in improving the accuracy and efficiency of cluster analysis in the face of high-dimensional data. Similar in-depth method of data analysis is contained in [8]-10].

In a study conducted in [11], researchers utilized the Gaussian Mixture Model (GMM) to analyze and quantify the inherent properties of peptides. Their findings demonstrated a superior performance compared to the complex Matrix Assisted Laser Desorption Ionization Time-of-Flight (MALDI-TOF) approach. This same approach was used to detect glaucomatous progression.

Furthermore, in another study referenced as [12], a combination of GMM and Support Vector Regression (SVR) was employed to predict the degradation of Aviation Piston Pump Performance. The results indicated that the joint GMM and SVR methodology outperformed the use of SVR alone.

Additionally, a Cooperative Localization-based technique utilizing the GMM in Wireless Sensor Networks (WSNs) was proposed in [13]. The authors demonstrated that this technique was more resilient against outliers compared to traditional least square algorithms.

Furthermore, a quantitative comparison between our proposed GMM with the firm iterative EM algorithm and the commonly used Gaussian probability distribution analysis method [2, 14-16] is provided using the Akaike Information Criterion (AIC).

In [17], the authors proposed a GMM with a robust iterative Expectation-Maximization (EM) algorithm that efficiently determines mixture model components and parameters for improved analysis and modeling of signal power coverage data.

In this paper, a Gaussian Mixture Model (GMM) with exhaustive cluster learning algorithm, emphasizing the significance of thorough GMM standardization is proposed and engaged in analyzing and deriving optimal insights from complex 5G-NR SINR datasets. Through a comprehensive methodology and implementation of the tuning algorithm, this study aims to provide valuable insights for enhancing cluster learning processes in the realm of 5G network performance analysis. By representing data points as a mixture of Gaussian distributions, GMM enables the identification of underlying patterns and structures within complex datasets.

3. Methodology

Description of 5G-NR SINR Dataset Collection Method

We explore the Cellular-Z software to collect SINR (Signal-to-Interference-plus-Noise Ratio) data, which is a crucial metric for evaluating the quality of 5G-NR cellular network connection. The real-time measured SINR data was performed in a typical urban environment in the United Kingdom. This data indicates the strength of the desired signal relative to interference and noise, influencing data transfer speed and reliability. Cellular-Z is a mobile app that can be used to monitor various signal strength parameters, including SINR, RSRP (Received Signal Power), and RSRQ (Received Signal Quality).

The collected SINR data can be used to assess the overall network quality and identify potential areas for improvement. For example, low SINR values may indicate interference issues or weak signal strength.

The Gaussian Mixture Model and Description of its Exhaustive Tuning Approach

The GMM is a distinctive machine learning based model for data clustering [17]. GMM clustering can accommodate clusters that have different sizes and correlation structures within them. Like many clustering techniques, GMM clustering provides means to specify the number of clusters before fitting the model. The number of clusters specifies the number of components in the GMM.

Choosing the appropriate clustering algorithm for the GMM is crucial for effective cluster learning[17]. In this paper, we adopt the exhaustive clustering GMM algorithm.The selection of GMM for 5G-NR SINR Quality datasets is justified by its ability to model complex data distributions and identify latent clusters with Gaussian characteristics, aligning well with the nature of SINR data.

Let x, k and Σ indicate the acquired SINR data set, desired components (i.e., clusters) number and the entire components covariance structure. The desired components number, k in a GMM determines the number of subpopulations, or clusters.

In Matlab computational environment, the following stepwise methods is adopted to engage the GMM clustering:

The exhaustive clustering GMM algorithm

i. Choose the appropriate clustering algorithm

ii. Load the data set x into the Matlab2024b workspace:

```
load('dataset ('x'));

SINR[dB]= [ ]';
```

iii. Select a (k,Σ) pair

iv. Engage the GMM with the selected (k,Σ) pair to fit the specific data set, x.

v. Specify GMM component number and 1000 maximum iterations for the selected algorithm:

```
k = 3; GMM component number
options = statset('MaxIter',1000);
```

vi. Specify the covariance matrices structure for all components-diagonal or full:

```
Sigma = {'diagonal','full'}; Options for covariance matrix structure
nSigma = numel(Sigma);
```

vii. Fit a GMM model to the datasets with the k components and covariance structure option:

```
gmfit = fitgmdist(X,k,'CovarianceType',Sigma{i}, …); Fitted GMM
clusterX = cluster(gmfit,X); Cluster index
mahalDist = mahal(gmfit,X0); each grid point distance to the GMM component
```

viii. Draw ellipsoids over each GMM component and show clustering result

ix. Examine the fitting performance using BIC and AIC indicators.

x. Repeat steps i - iv until the selected (k,Σ) pairs are exhausted.

xi. Document the resultant fitted GMM that stabilities and provide the lowest AIC and BIC values.

xii.visualize the results in 2-D

4. Results and Analysis

Shown in Figs. 1-5 are the graphical results of the SINR clustering based learning distribution results obtained using the proposed exhaustive GMM algorithm. As pointed out in previous sections, the SINR stands for Signal-to-Interference-plus-Noise Ratio) and its values ranges typically range from 0 to 35db across the measurement locations.A higher SINR value indicates better signal quality, with values above 20 dB generally considered excellent. Higher-order modulation schemes like 256QAM in 5G require higher SINR values for reliable performance compared to lower-order schemes like 16QAM. The field tests show that the SINR distributions with most areas are having values between 5 and 15 dB, which is sufficient for 16QAM and lower-order modulation in 64QAM. Higher SINR values directly translate to higher data speeds and a more reliable connection. Values below 0 dB indicate that the received signal is weaker than the noise and interference, leading to potential connection drops and very low data speeds.This could imply that there is more noise in the received signal than the useful part, and the probability of losing a connection certainly possible at the SINR quality level.

More importantly, Figs. 1-5 also reveal the cluster distribution impact with different k components fitted to the measured SINR data. Different k components, ranging from 1-6 were used for the cluster distribution to ascertain which one best provide the best cluster fitting to the measured data. Each component is defined by its mean and covariance.

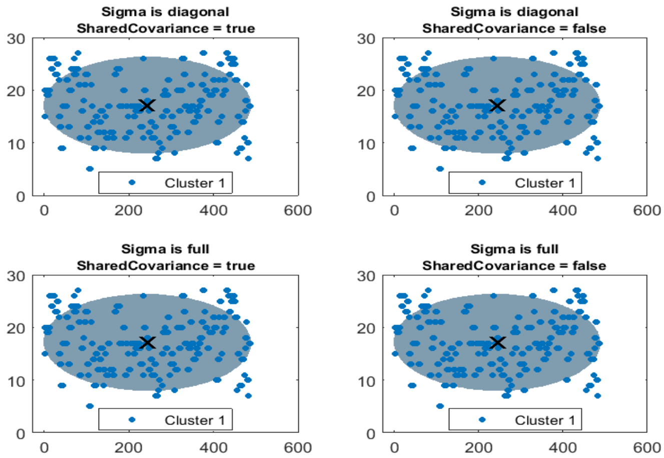

Fig. 1: Learned SINR Data using GMM model with proposed Algorithm, k=1

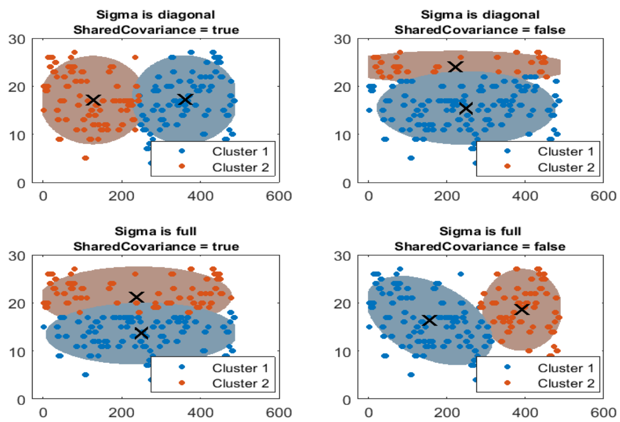

Fig. 2: Learned SINR Data using GMM model with proposed Algorithm, k=2

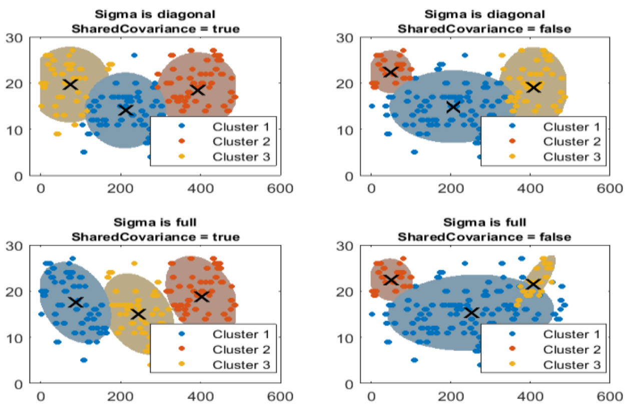

Fig. 3: Learned SINR Data using GMM model with proposed Algorithm, k=3

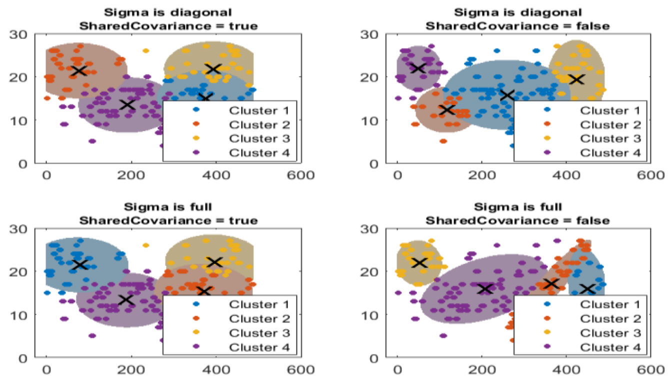

Fig. 4: Learned SINR Data using GMM model with proposed Algorithm, k=4

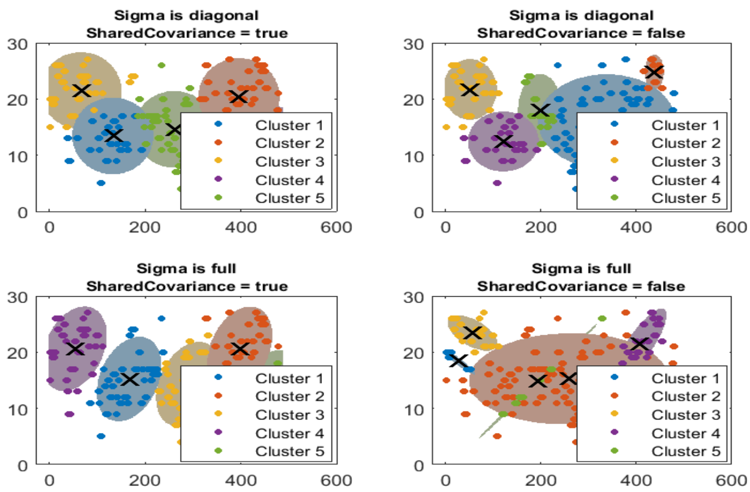

Fig.5: Learned SINR Data using GMM model with proposed Algorithm, k=5

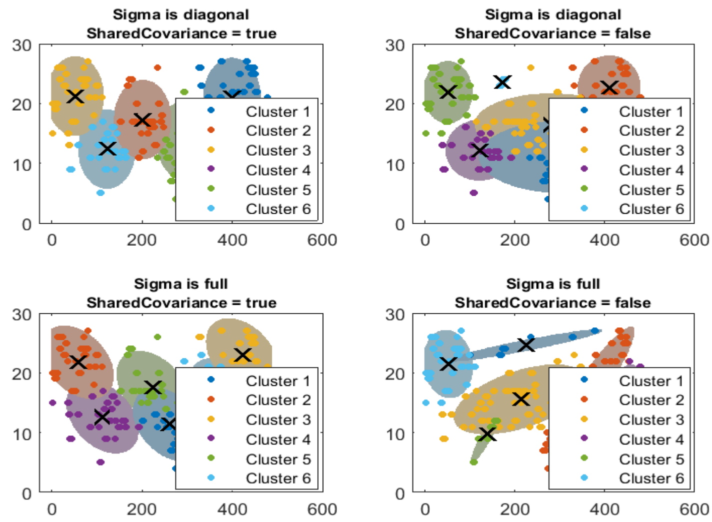

Fig.6: Learned SINR Data using GMM model with proposed Algorithm, k=6

By means of Akaike Information Criterion (AIC) and Bayesian information criteria (BIC) statistical analysis, Figs. 7-12 are presented to help choose the best fitting GMM k components model to the SINR data. Both the

AIC and BIC are likelihood-based measures of model fit at different complexity penalty.Lower AIC or BIC values indicate better fitting models. According to the plotted computed AIC and BIC values, the best model cluster has 4 components and a full, shared covariance matrix structure. Thus, the proposed exhaustive algorithm reveals that the GMM with 4 components with a full, shared covariance matrix structure provide the cluster fitting to the targeted measured 5G-NR SINR data.

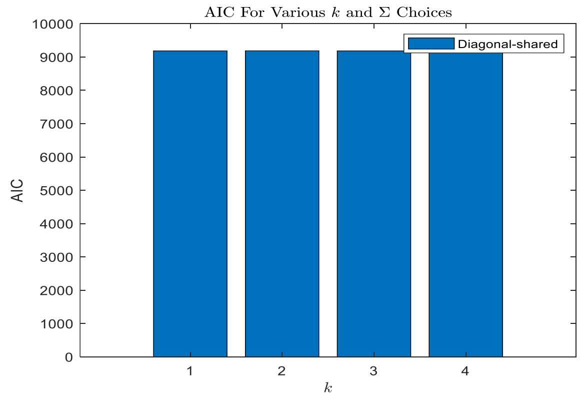

Fig. 7: GMM model cluster based fitting performance with AIC for k=1

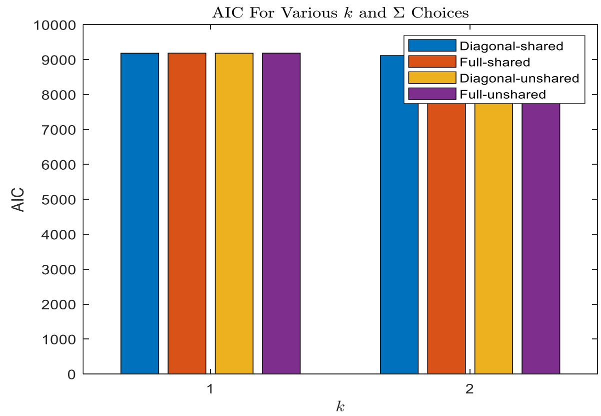

Fig. 8: GMM model cluster based fitting performance with AIC for k=2

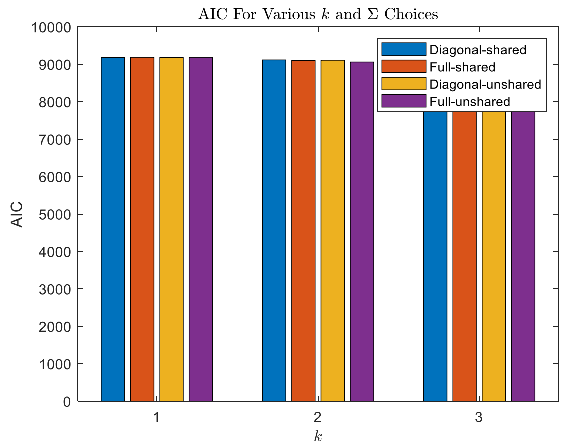

Fig. 9: GMM model cluster based fitting performance with AIC for k=3

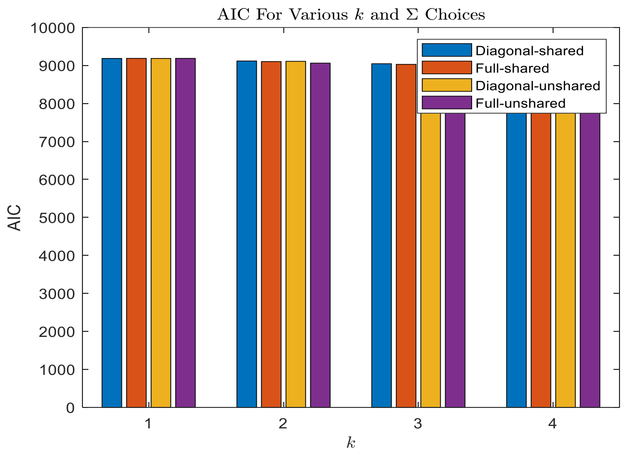

Fig. 10: GMM model cluster based fitting performance with AIC for k=4

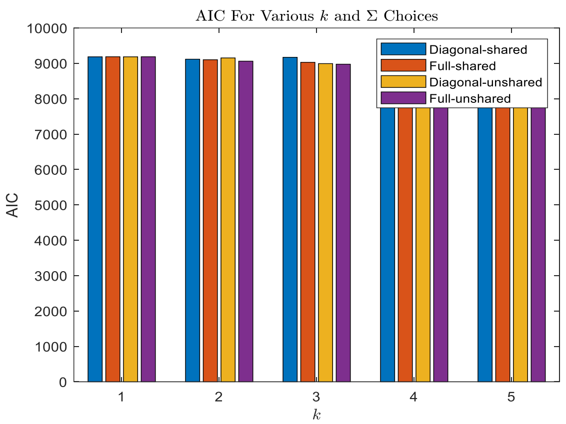

Fig. 11: GMM model cluster based fitting performance with AIC for k=5

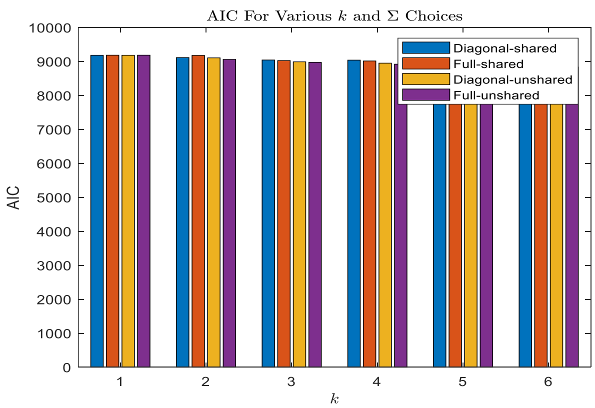

Fig. 12: GMM model cluster based fitting performance with AIC for k=6

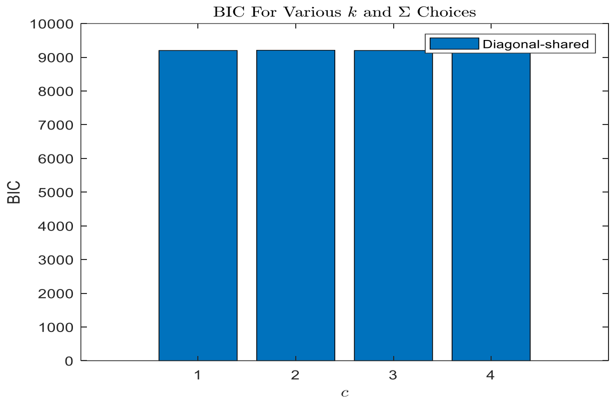

Fig. 13: GMM model cluster based fitting performance with BIC for k=1

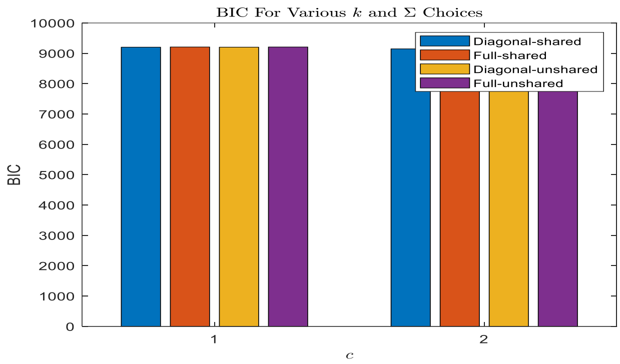

Fig. 14: GMM model cluster based fitting performance with BIC for k=2

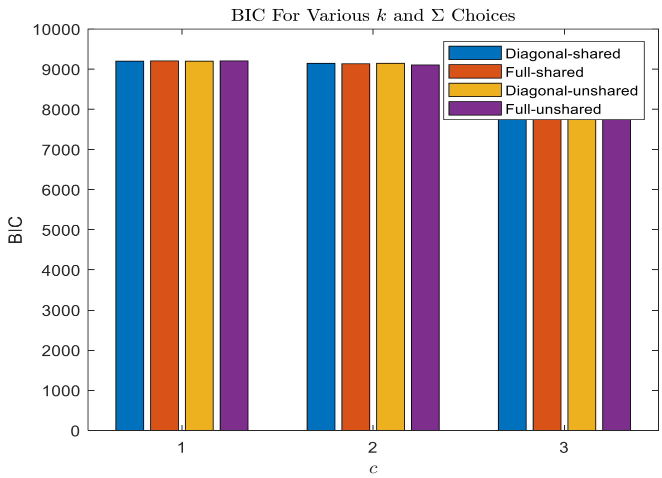

Fig. 15: GMM model cluster based fitting performance with BIC for k=3

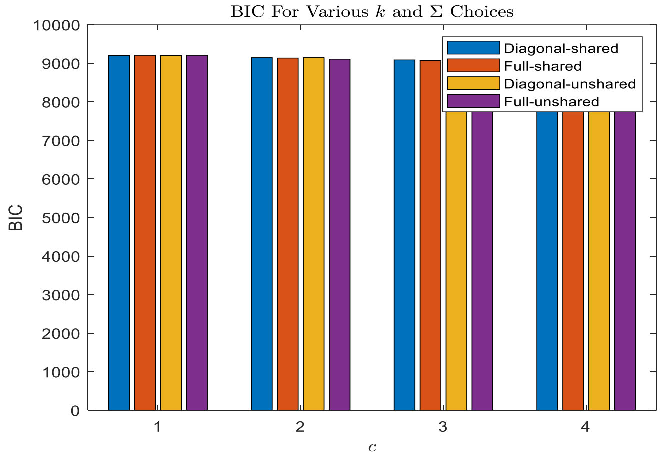

Fig.16: GMM model cluster based fitting performance with BIC for k=4

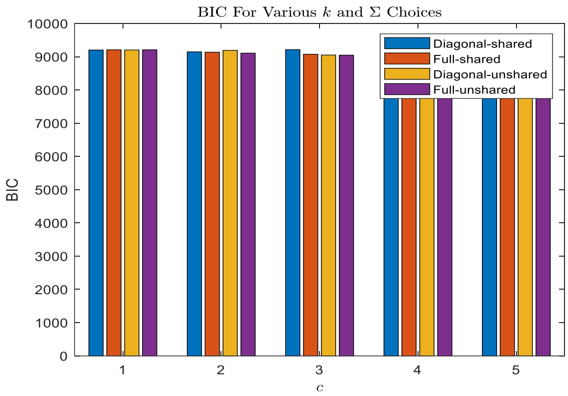

Fig. 17: GMM model cluster based fitting performance with BIC for k=5

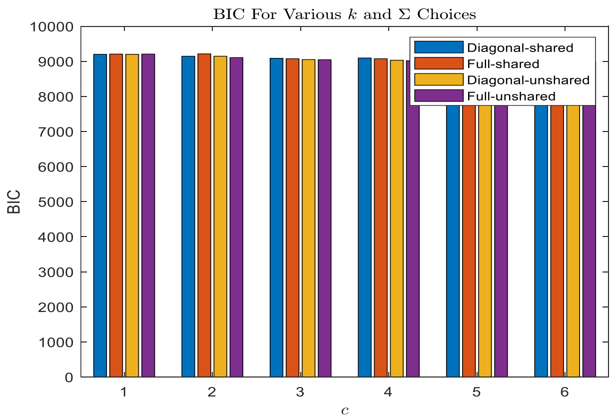

Fig. 18: GMM model cluster based fitting performance with BIC for k=6

5. **Conclusion**

By means of exhaustive tuning algorithm, this paper shows how to determine the best Gaussian mixture model (GMM) fit by adjusting the number of components and the component covariance matrix structure. Particularly, we paper showcased the the efficacy in optimizing cluster learning for 5G-NR SINR quality datasets. By shedding light on the importance of fine-tuning parameters and implementing advanced methodologies, this study contributes to the advancement of data analysis techniques in the realm of 5G network optimization. As we continue to explore and innovate in the field of wireless communication technologies, the insights gained from this research pave the way for further enhancements in cluster learning methodologies for 5G networks, ultimately leading to improved network performance and user experience.

References

1. A. Rezaei, P. Schulz, R. S. Ganesan, A. Awada, I. Viering and G. Fettweis, "Efficient Approximation of SINR and Throughput in 5G NR via Sparsity and Interference Aggregation," *2023 IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Toronto, ON, Canada, 2023, pp. 1-7, doi: 10.1109/PIMRC56721.2023.10293808.

2. Isabona, J and Osaigbovo, I.A.: Investigating predictive capabilities of RBFNN, MLPNN and GRNN models for LTE cellular network radio signal power datasets, FUOYE. J. Eng. Technol.4(1), 155–159 (2017)

3. V.C Ebhota, J. Isabona, and V.M Srivastava, Base line knowledge on propagation modelling and prediction techniques in wireless communication networks, Journal of Engineering and Applied Sciences 13 (7), 1919-1934

4. J. Isabona, and D.O Divine, Adaptation of Propagation Model Parameters toward Efficient Cellular Network Planning using Robust LAD Algorithm, International Journal Wireless and Microwave Technologies 10, 13-24.

5. Torre F.D.L and Kanade T 2006 Discriminative cluster analysis In Theory and Novel Applications of Machine Learning

6.Xie J, Girshick R and Farhadi A 2015 Unsupervised deep embedding for clustering analysis Computer Science Gaussian Mixture Model (GMM) for Cluster Learning.

7. Xianghong Lin, Xiaofei Yang and Ying Li, A Deep Clustering Algorithm based on Gaussian Mixture Model, IOP Conf. Series: Journal of Physics: Conf. Series 1302 (2019) doi:10.1088/1742-6596/1302/3/032012

8. J Isabona, Maximising coverage and capacity with QoS Guarantee in GSM network by means of cell cluster optimization, International Journal of Advanced Research in Physical Science, vol. 1 (6), 44-55

9. M. Ekpenyong, J. Isabona and E. Isong, "Handoffs Decision Optimization of Mobile Cellular Networks," 2015 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 2015, pp. 697-702, doi: 10.1109/CSCI.2015.155.

10. Isabona, J., Srivastava, V.M. Downlink Massive MIMO Systems: Achievable Sum Rates and Energy Efficiency Perspective for Future 5G Systems. *Wireless Pers Commun* 96, 2779–2796 (2017). https://doi.org/10.1007/s11277-017-4324-y

11. Spainhour, J.C.G., Janech, M.G., Schwacke, J.H., Velez, J.C.Q.: Ramakrishnan V the application of gaussian mixture models for signal quantification in MALDI-ToF mass spectrometry of peptides. PLoS ONE 9(11), e111016 (2014). https://doi.org/10.1371/journal.pone.011101612.

12. Lu, C., Wang, S.: Performance degradation prediction based on a Gaussian mixture model and optimized support vector regression for an aviation piston pump. Sensors 20, 3854 (2020).https://doi.org/10.3390/s2014385413.

13. Yin, F., Fritsche, C., Jin, D., Gustafsson, F., Zoubir, A.M.: cooperative localization in WSNs using Gaussian mixture modeling: distributed ECM algorithms. IEEE Trans. Signal Process.63(6), 1448–1463 (2015)14.

14. Abiodun, C.I., Ojo, J.S.: Determination of probability distribution function for modelling path loss for wireless channels applications over micro-cellular environments of Ondo State,Southwestern Nigeria. World Sci. News 118, 74–88 (2019)15.

15. Isabona, J.: Maximum likelihood parameter based estimation for in-depth prognosis investigation of stochastic electric field strength data. BIU J. Basic Appl. Sci. 4(1), 127–136(2019)

16. Isabona, J., Konyeha, C.C.: Experimental study of UMTS Radio signal propagation characteristics by field measurement. Am. J. Eng. Res. 2(2), 99–106 (2013)

17. Joseph Isabona and Divine Osamiromwen Ojuh, A Gaussian Mixture Model with Firm Expectation-Maximization Algorithm for Effective Signal Power Coverage Estimation, In book: Information and Communication Technology and Applications, Third International Conference, ICTA 2020, Minna, Nigeria, November 24–27, 2020, Revised Selected Papers, DOI: 10.1007/978-3-030-69143-1_8