



PIGMES: Partial Incremental Gaussian Mixture Model with Efficient Storage

Meriem Bouzouad, Yasmine Benhamadi, Camélia Slimani, Mouloud Koudil,
Jalil Boukhobza

► To cite this version:

Meriem Bouzouad, Yasmine Benhamadi, Camélia Slimani, Mouloud Koudil, Jalil Boukhobza. PIGMMES: Partial Incremental Gaussian Mixture Model with Efficient Storage. Per3S - Performance and Scalability of Storage Systems, May 2023, Paris, France. hal-04303907

HAL Id: hal-04303907

<https://hal.science/hal-04303907>

Submitted on 23 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PIGMMEs: Partial Incremental Gaussian Mixture Model with Efficient Storage

Meriem Bouzouad^{*1,2}, Yasmine Benhamadi^{*1,2}, Camélia Slimani¹,
Mouloud Koudil^{2,3}, Jalil Boukhobza¹

¹ ENSTA Bretagne, Lab-STICC, CNRS, UMR 6285, F-29200 Brest, France

² Ecole nationale Supérieure d'Informatique (ESI-Alger)

³ Laboratoire des Méthodes de Conception des Systèmes (LMCS), Alger, Algérie

* These 2 authors contributed equally to the design of this contribution



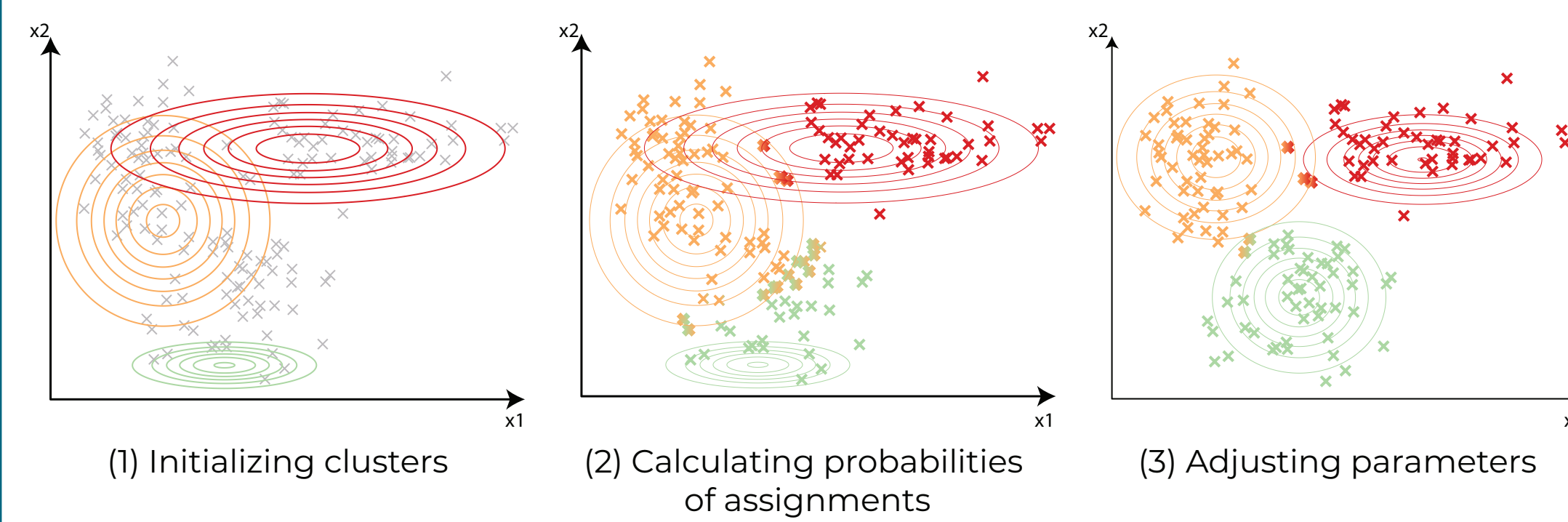
1- Introduction

- Edge intelligence motivates embedded machine learning optimization [1].
- Strict memory constraints impact machine learning (ML) algorithms' performance because of I/O operations [2][3][4].
- Gaussian mixture model (GMM) is a popular ML algorithm used in several embedded applications, e.g. speaker identification [5], sensor data analytics [6], healthcare [7]..etc.

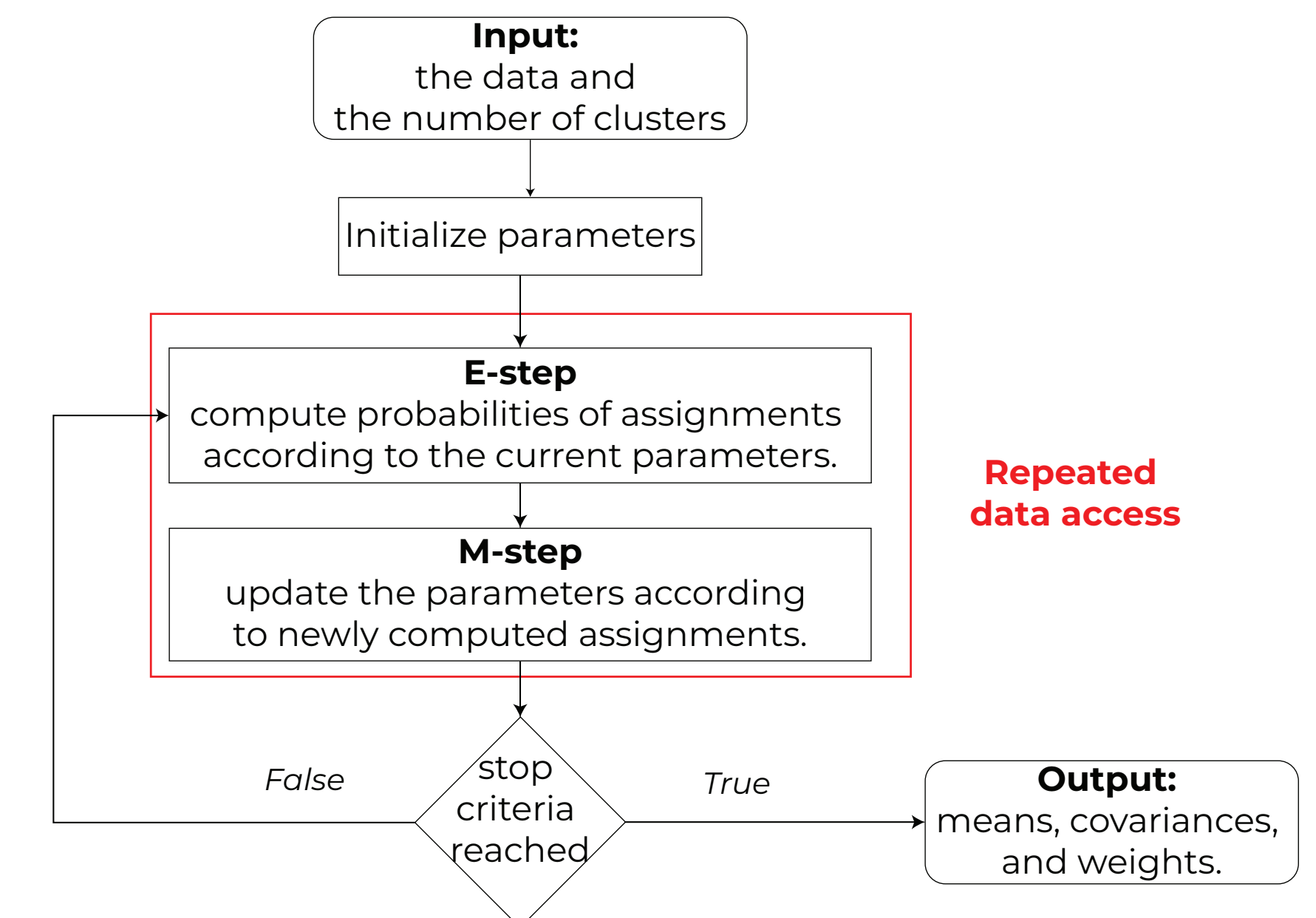
Problem statement: how to optimize the GMMs training algorithm on memory-constrained embedded systems?

2- Gaussian Mixture Models for clustering

In the context of clustering, GMMs model each cluster by a Gaussian density (mean and covariance matrix), as well as the probabilities of assignment to each cluster (weights).



GMM bivariate clustering example

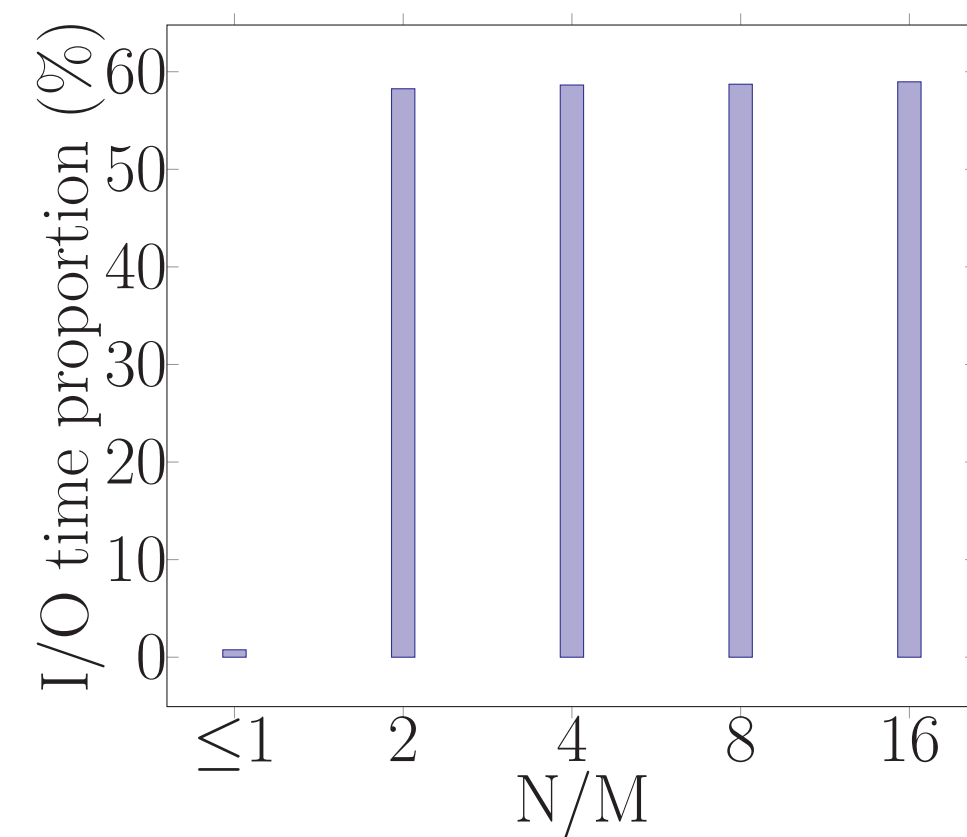


The Expectation-Maximization (EM) algorithm for GMM learning

3- Motivation

N : Number of observations in the data-set.

M : Number of observations that can be contained in memory.



Proportion of time spend on I/O operations for different values of N/M

4- PIGMMES: Partial Incremental Gaussian Mixture Model with Efficient Storage

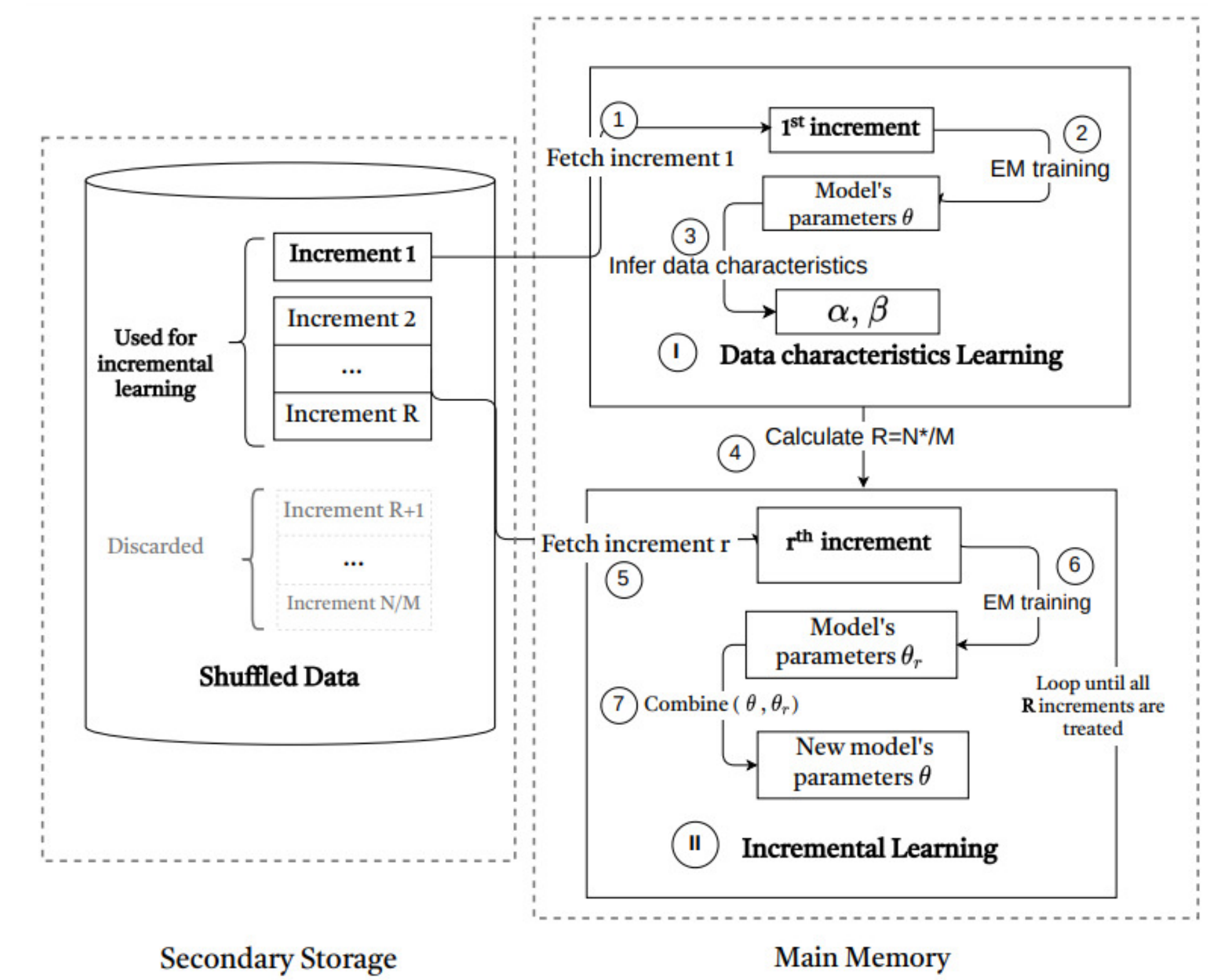
PIGMMES is based on two fundamental ideas:

- Restricting the volume of the training set using inferred data characteristics from the first increment (1).
- Doing an incremental learning of GMM on the restricted dataset to reduce I/Os cost.

$$N^* = M + \frac{[(1 - \beta) + \alpha](N - M)}{2} \quad (1)$$

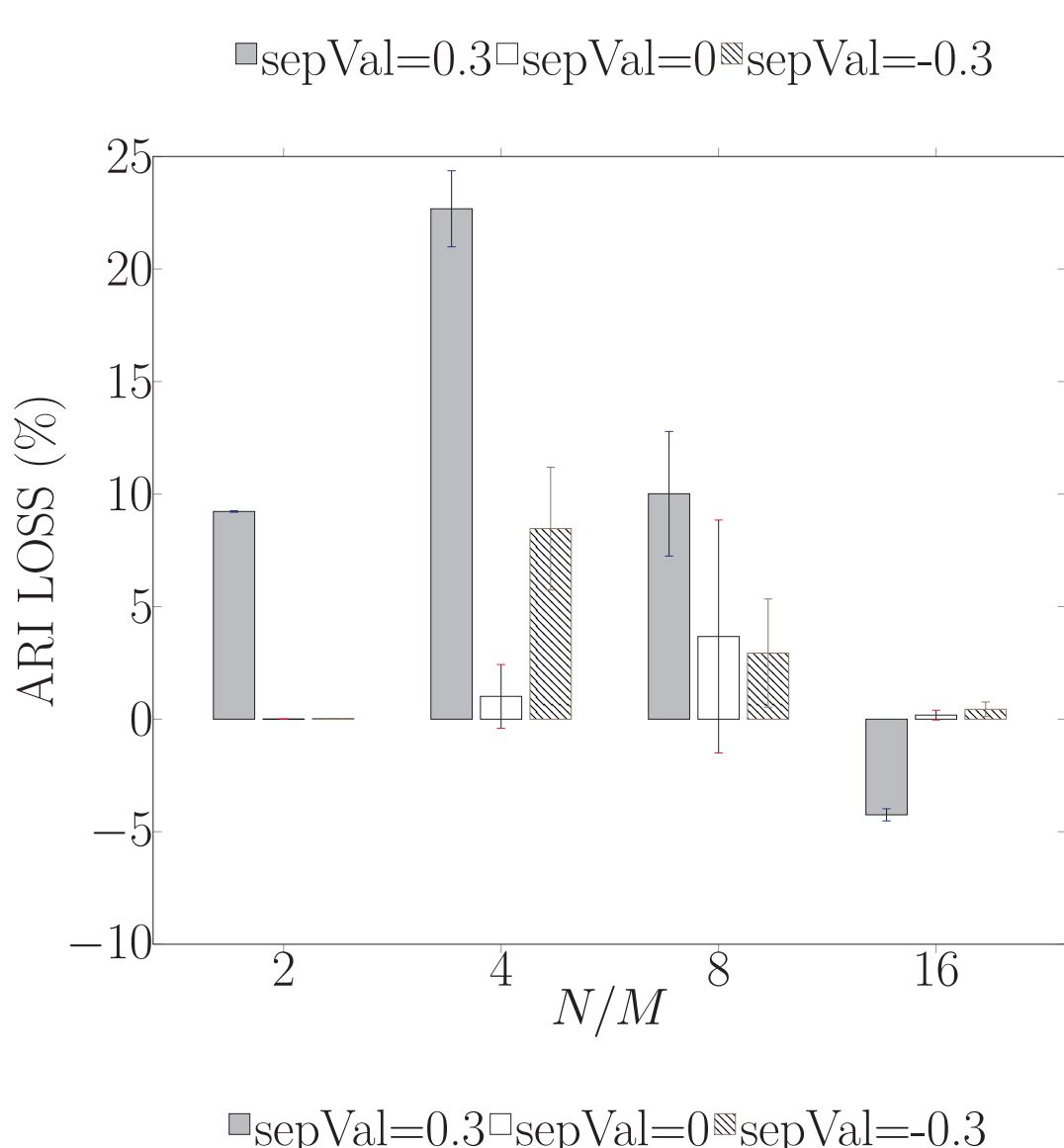
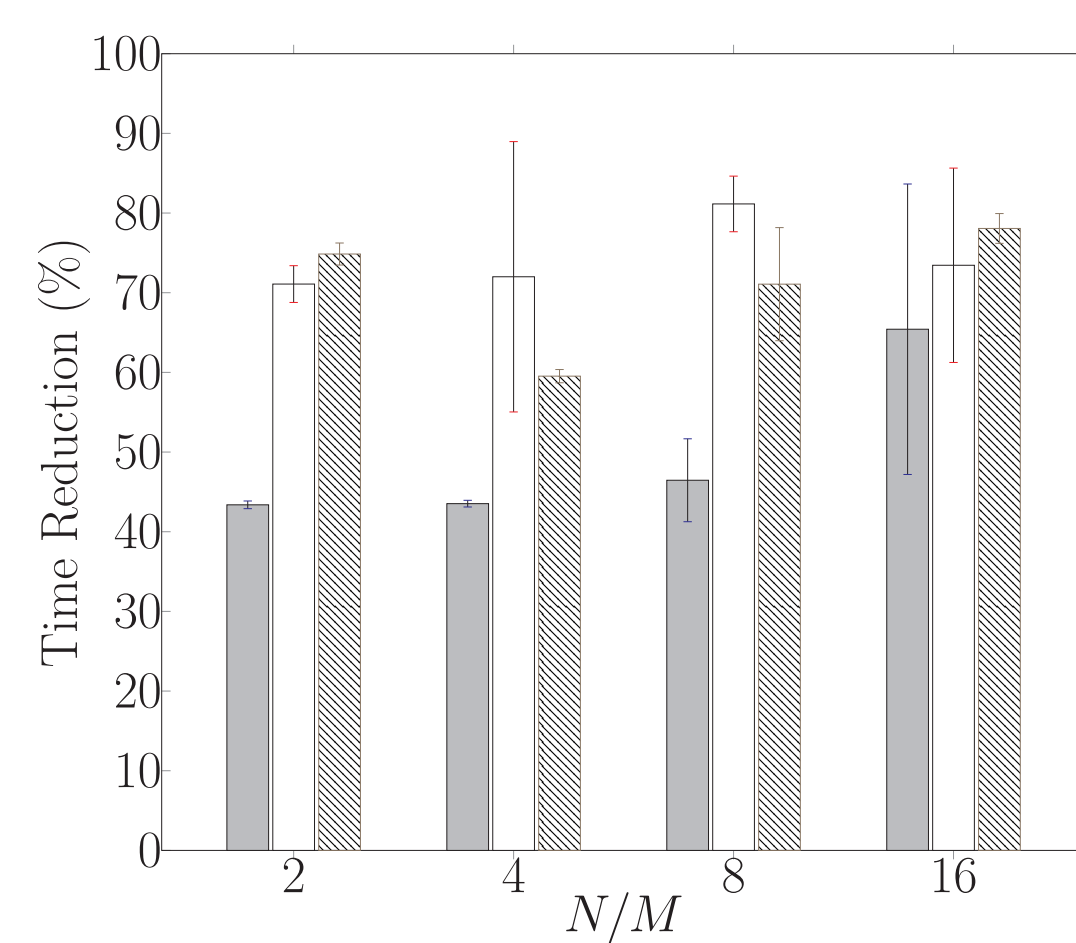
$\alpha \in [0, 1]$ characterizes the complexity of the data.

$\beta \in [0, 1]$ characterizes the balance of the clusters.



5- Evaluation

- The method was evaluated on synthetic datasets of 100MB each, with different separation values, using the Odroid XU4Q platform [8].
- The quality of clustering of the proposed method was compared to that of the traditional EM algorithm using the Adjusted Rand Index (ARI).



6- Related work

- [9] and [10] propose incremental online GMM for data stream applications with different merging strategies.
- [11] propose a sub-sampling approach for GMMs that determines the size of a uniform subsample and then generalizes the results through additional iterations.
- [12] propose a sub-sampling approach that consists in selecting a representative subset of the data using K-means.

We borrowed and upgraded ideas from both incremental and sub-sampling GMMs to elaborate PIGMMES.

7- Conclusion and Perspectives

- Our I/O optimization of the GMM training algorithm is based on two key contributions: (1) an adaptive uniform sub-sampling approach and (2) a divide-and-conquer learning strategy.
- Our method achieves, on average, a 55% improvement in overall execution time.
- For future work, we are aiming to:
 1. Reduce our method's sensitivity to initialization.
 2. Generalize to an undetermined number of clusters.
 3. Implement a parallel and/or a distributed version of the method.

References

- [1] Massimo Merenda et al. Edge machine learning for ai-enabled iot devices: A review. *Sensors*, 20(9):2533, 2020.
- [2] Camélia Slimani et al. K-mlio: Enabling k-means for large data-sets and memory constrained embedded systems. In *2019 IEEE MASCOTS*, 2019.
- [3] Camélia Slimani et al. Rafio: A random forest i/o-aware algorithm. In *Proceedings of the 36th Annual ACM SAC*, page 521–528. ACM, 2021.
- [4] Camélia Slimani et al. Accelerating random forest on memory-constrained devices through data storage optimization. *IEEE TC*, 72, 2023.
- [5] Alberto Gianelli et al. Low power speaker identification using look up-free gaussian mixture model in cmos. In *2019 IEEE COOL CHIPS*, 2019.
- [6] Jongmin Lee et al. Integrating machine learning in embedded sensor systems for internet-of-things applications. In *2016 IEEE ISSPIT*, pages 290–294, 2016.
- [7] Vassilis Alimisis et al. Nanopower integrated gaussian mixture model classifier for epileptic seizure prediction. *Bioengineering*, 9, 2022.
- [8] Rob Roy and Venkat Bommakanti. *Odroid XU4 - User Manuel*. Hard Kernel, Ltd, 2015.
- [9] Mingzhou Song et al. Highly efficient incremental estimation of gaussian mixture models for online data stream clustering. volume 5803, pages 174–183. SPIE, 2005.
- [10] Juan M Acevedo-Valle et al. Multivariate regression with incremental learning of gaussian mixture models. In *CCIA*, pages 196–205, 2017.
- [11] Ron Wehrens et al. Model-based clustering for image segmentation and large datasets via sampling. *Journal of Classification*.
- [12] Mario Lucic et al. Training gaussian mixture models at scale via coresets. *The Journal of Machine Learning Research*, 18(1):5885–5909, 2017.