Master thesis proposal: Searching for the Best Clustering Assignment from Training Dynamics

Mamdouh Aljoud

November 2025

1 Introduction

Deep clustering has emerged as a powerful approach to unsupervised learning by combining the strengths of deep representation learning with traditional clustering. By jointly optimizing a neural network encoder and a clustering objective, methods such as DEC, IDEC, DCN, and DKM can learn feature representations that reveal the underlying structure of complex, high-dimensional data. Unlike conventional clustering algorithms that operate directly on raw features, deep clustering models can capture nonlinear relationships and semantic similarities in the data. However, these methods also inherit the instability common to deep learning models. The quality of the cluster assignments often fluctuates during training, sometimes reaching its peak well before full convergence. Identifying the point at which the best clustering occurs usually requires labeled data, which contradicts the goal of unsupervised learning. This thesis explores whether the training dynamics—in particular, the loss curve and the embeddingsproduced at each epoch—can be used to automatically detect the best clustering assignment generated during optimization, without relying on ground-truth labels.

2 Problem Statement

Can we identify the best cluster assignments produced by the DC algorithm based on the loss curve and the embeddings at each epoch?

3 Tasks

- Collect runs from multiple datasets and deep clustering algorithms (DEC, IDEC, DCN, DKM)
- Survey methods for assessing representation quality without labels
- Review time-series analysis and forecasting techniques for modeling loss curves

- Extract loss-curve and representation-based features per epoch.
- Build a label-free clustering quality assessor
- Compare the selected epoch's clustering quality to the final epoch using true metrics (ACC, NMI, ARI)
- Document and evaluate generalization across datasets and models

4 Requirements

- Strong background in deep learning and machine learning
- Good coding skills (preferably PyTorch)

5 Contact

If you are interested send an email to aljoud@dbs.ifi.lmu.de with your CV and transcript and highlight your interest in this topic.