Bachelor thesis proposal: A Systematic Study of Training Stability in Deep Clustering Algorithms

Mamdouh Aljoud

November 2025

1 Introduction

Deep clustering combines neural networks with clustering methods to learn useful representations and group data without labels. Algorithms like DEC, IDEC, DCN, and DKMjointly learn feature embeddings and cluster assignments, allowing them to handle complex, high-dimensional data where traditional clustering often fails. These methods have shown promising results across different domains, from images to text and tabular data. Still, deep clustering methods often suffer from instability. Small changes in initialization, architecture, or hyperparameters can lead to very different clustering results. This makes it difficult to compare models fairly or to know which method can be trusted in practice. This thesis will take a closer look at this issue by studying how stable different deep clustering algorithms are and whether some perform more consistently than others across different datasets and data modalities.

2 Problem Statement

Which deep clustering algorithms are the most stable across runs and datasets, and does their stability depend on the data modality?

3 Tasks

- Train multiple deep clustering algorithms (DEC, IDEC, DCN, DKM) on different datasets and data modalities.
- Repeat experiments with different random seeds and hyperparameters.
- Measure variability in results using metrics like ACC, NMI, and ARI.
- Analyze and visualize stability across algorithms and datasets.
- Identify which models are most robust and where they fail.

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.