Bachelor thesis proposal: Understanding the Impact of Hyperparameters on Deep Clustering Performance

Mamdouh Aljoud

November 2025

1 Introduction

Deep clustering combines neural networks with clustering methods to learn meaningful data representations and group samples without supervision. By jointly optimizing feature learning and clustering, algorithms such as DEC, IDEC, DCN, and DKM can discover structure in complex, high-dimensional data where traditional clustering methods often fail. However, these methods rely on many hyperparameters — such as the learning rate, number of clusters, bottleneck size, or the weighting between different loss terms — that strongly influence the outcome. Small changes in these values can lead to large variations in clustering quality and convergence behavior. As a result, understanding how each hyperparameter affects performance is essential for using these models reliably and efficiently. This thesis aims to systematically study how different hyperparameters influence the performance of deep clustering algorithms and to identify which ones have the greatest impact on results.

2 Problem Statement

Which hyperparameters have the strongest effect on the performance and stability of deep clustering algorithms?

3 Tasks

- Select a few representative deep clustering models (DEC, IDEC, DCN, DKM).
- Define key hyperparameters to vary (e.g. learning rate, number of clusters, hidden size, regularization).
- Run controlled experiments on several datasets.

- Record performance metrics (ACC, NMI, ARI) for each configuration.
- Use statistical analysis to identify which hyperparameters matter most

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.