

Fast and Automatic Subspace Clustering

Applying Modern Clustering Methods to the Subspace Clustering Problem

Yan Zhu, Master of Science in Computer Science
University of Dublin, Trinity College, 2022

Supervisor: Mimi Zhang, Joshua Tobin

Due to the continuous development of big data technologies, researchers have an increasing demand for data analysis of high-dimensional data. Clustering analysis, one of the critical data analysis techniques, has many applications. For example, in bioinformatics or NLP, input datasets include tens to thousands of features (or dimensions), which makes data processing quite difficult. Clustering methods can help researchers to extract data with critical features from large data sets. Most modern subspace clustering methods apply spectral clustering algorithms for the final clustering step, which needs the exact number of clusters in advance. This thesis aims to optimise the final segmentation step to improve the algorithm's stability and obtain a method with good performance and good clustering results suitable for clustering analysis of high-dimensional data.