1. k-means [17], along with many conventional clustering algorithms such as Gaussian Mixture Models (GMMs) [23], DBSCAN [4], and hierarchical algorithms [12] typically require hand engineered features to be created for each dataset and task. Further, these features may then be analysed using another process, feature selection, in order to eliminate redundant or poor quality features. This task is even more challenging in the unsupervised setting. Additionally, it is a time-consuming and brittle process, with the choice of features having a large influence over the subsequent performance of the clustering algorithm.
2. Recent advances in deep learning have paved the way for algorithms which can effectively learn from raw data, bypassing the need for manual feature extraction and selection.
3. Deep clustering refers to the process of clustering with deep neural networks,
4. We propose a simple approach, N2D, that effectively replaces the clustering network with a manifold learning technique on top of the autoencoded representation.
5. Given this updated embedding, we can then cluster it with conventional nondeep clustering algorithms. By doing so, N2D replaces the complexity of the clustering network with a manifold learning method and straightforward non-deep clustering algorithm,
6. One important question is which manifold learning technique to apply to the autoencoded representation.
7. More recently, UMAP [19] has been proposed, which while also local, has been shown to better preserve global structure. All of these methods seek to utilize the distances between points in order to better learn the underlying structure, and we posit that they will improve the clusterability of an autoencoded embedding.
8. we propose a framework, N2D, where in contrast to recent deep clustering techniques, we replace the deep clustering network with a manifold learning method, and shallow cluster the resulting re-embedded space.
9. Clustering algorithms can be broadly categorized into two different categories, hierarchical clustering and partitional clustering.
10. Deep clustering methods use deep neural networks to cluster, typically involving two different processes, one where a representation is learned, and one where the actual clustering occurs. This process may occur separately or jointly.