Categorization of images without labels

*

Himanshu Singh *2022217*
Jaleel Ahmed Radhu Khawaja *2022225*

I. MOTIVATION

Can we automatically group images into semantically meaningful clusters when ground-truth annotations are absent? The problem at hand is unsupervised clustering, specifically clustering without labelled data. The objective is to group similar instances without prior knowledge of their class labels. This is a common scenario in many real-world applications where obtaining labelled data is expensive or impractical. This project aims to explore and develop techniques that can automatically identify patterns and group data points based solely on their intrinsic properties.

II. LITERATURE REVIEW

A. SCAN: Learning to Classify Images without Labels

Many works have been done on the following, which we are trying to solve. Following are two papers we referred to most [1] SCAN: Learning to Classify Images without Labels This paper does the task in two steps. In the first step, they learn feature representations through a pretext task. They propose to mine the nearest neighbours of each image based on feature similarity. They empirically found that, in most cases, these nearest neighbours belong to the same semantic class, rendering them appropriate for semantic clustering. In the second step, they integrate the semantically meaningful nearest neighbour as a prior into a learnable algorithm. It then tries to classify the images and their nearest neighbour, found in the first step, by maximising the dot product after softmax. Which makes the network produce consistent and discriminative (one-hot) predictions. Consistency is gained by taking it as a dot-product, which makes the distribution less variant.

B. Deep Embedded K-Means Clustering

This method uses a similar idea where they start with the belief that high representational learning and clustering can reinforce each other. In this paper, they propose a method called DEKM (Deep Embedded K-Means). They use an auto-encoder to get the feature representation in a lower dimension with all the high-level information. Since the auto-encoder-generated embedding space may have no prominent cluster structures; they transform the embedding space to a new space that reveals the cluster-structure information. This is achieved by an orthonormal transformation matrix, which contains the eigenvectors of the within-class scatter matrix of K-means. The eigenvalues indicate the importance of the eigenvectors' contributions to the cluster structure information in the new space. Their goal is to increase the cluster-structure information.

III. DATASET

- The MNIST dataset, a widely used benchmark dataset in machine learning, is employed for this project. MNIST consists of 28x28 pixel grayscale images of handwritten digits (0-9). It contains 60,000 training images and 10,000 testing images.
- The MNIST Fashion dataset includes 60,000 training and 10,000 testing images, each belonging to one of the following classes. T-shirt/top, trousers, Pullover, Dress, Coat, Sandal, Shirt, Sneaker, Bag and Ankle boot
- The CIFAR-10 dataset is a widely used benchmark dataset in computer vision and machine learning. The CIFAR-10 dataset comprises 60,000 32x32 pixel colour images in 10 classes, each representing a different object or category.

IV. PROPOSED ARCHITECTURE

First of all, for each dataset, since we are trying to classify the images without labels, apriori, we will not have any knowledge about the number of classes in the dataset; therefore, for this, we will have to rely on smart guesses will be based on past experiences and domain knowledge. Since we already know the number of classes for the datasets we are testing, we "guess" the number of classes equal to the actual number of classes. This also helps us ease the testing of our model. Initially, we started with exploring the traditional clustering algorithm like K-means and hierarchical clustering. They gave us a satisfactory performance of about 59 per cent and 66 per cent accuracy, respectively, on the MNIST dataset. After looking at the scatter plot of clusters formed by K-means and original labels by reducing its dimension to 2 using PCA, we saw that class entanglement was one major factor for such low accuracy. To solve this, we considered discarding the low-level information/noise (which we speculated caused the class entanglement) and extracting only the high-level information from the data. Thus, we tried embedding the data points in lower dimension space where high-level information is retained.

A. Embedding techniques:

- PCA: First, we started with the most basic linear transformation technique, i.e. PCA. It gave us the highest

Identify applicable funding agency here. If none, delete this.

performance of 57.3shown below). This didn't increase the accuracy rather decreased it. One reason we think of the lousy performance of PCA is that since PCA is a linear transformation. It can only capture linear relationships within the data and not non-linear ones. Hence, the problem of Data entanglement still needs to be solved.

- Auto-Encoders: Here, we first train the auto-encoder and then use the bottleneck layer of the auto-encoder to get the embedded space. Since the autoencoder uses a neural network, we can also capture non-linear relationships. We can get an accuracy of 63.5

To solve the problem of data entanglement, we also tried increasing the number of clusters. That's the value of K in K-means clustering. This gave us a good accuracy of 66clusters increased. The problem with increasing the number of clusters is that it is not consistent with the motivation and objectives of our project. Because it is obvious that as we increase the number of clusters, the accuracy is bound to increase. Thus, to solve this problem we are trying to look for a technique in which we try to increase the number of clusters and then merge them in a way such two or more clusters get assigned the same class. Further work needs to be done on this.

V. Results

	MNIST	MNIST-Fashion	CIFAR-10
K-means	59%	55%	21.9%
PCA+K-Means	57.3%	62%	22%
Agglomerative Clustering	65%	57%	-
AutoEncoder+K-Means	62%	TO DO	TO DO
AutoEncoder+ Agglomerative	73.52%	TO DO	TO DO

TABLE I
Performance comparison of clustering algorithms on Purity

	MNIST	MNIST-Fashion	CIFAR-10
K-means	0.5159	0.5249	0.07
PCA+K-Means	0.5199	0.5241	0.0751
Agglomerative Clustering	0.7113	0.6123	-
AutoEncoder+K-Means	0.3392	TO DO	TO DO
AutoEncoder+ Agglomerative	0.7501	TO DO	TO DO

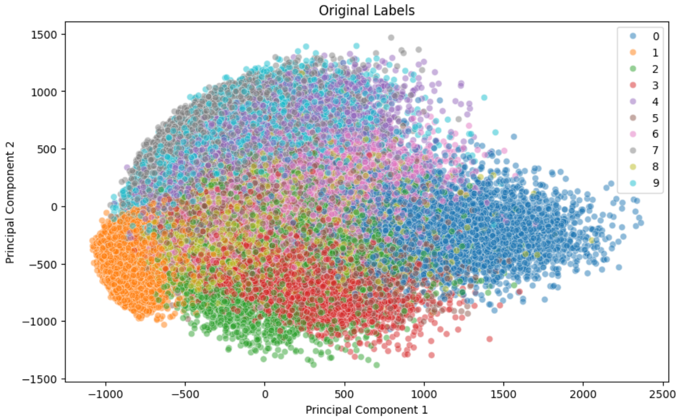
TABLE II
Performance comparison of clustering algorithms on NMI

VI. Visualisations

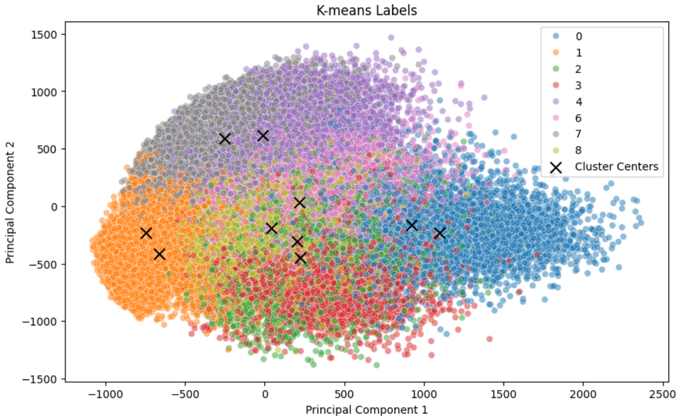
VII. Analysis

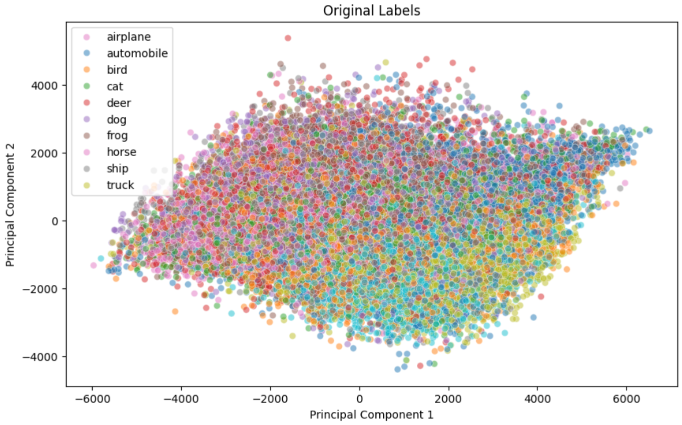
A. MNIST and Fashion MNIST Dataset

K-means alone achieved 59 percent accuracy, which is reasonable for a basic clustering algorithm without labels. PCA followed by K-means yielded slightly lower accuracy at 57.3 percent as all features were not included (max 100 were included), in the case of fashion mnist PCA performed better as in case of Fashion items, more pixels are required to show the data, so reducing dimensionality does not result in losing a lot of information. Using AutoEncoders for embedding improved accuracy to 62 percent, showing promise in capturing non-linear relationships within the data. Using AutoEncoders Agglomerative Clustering with autoencoders further improved the result as it also performed well with raw data.

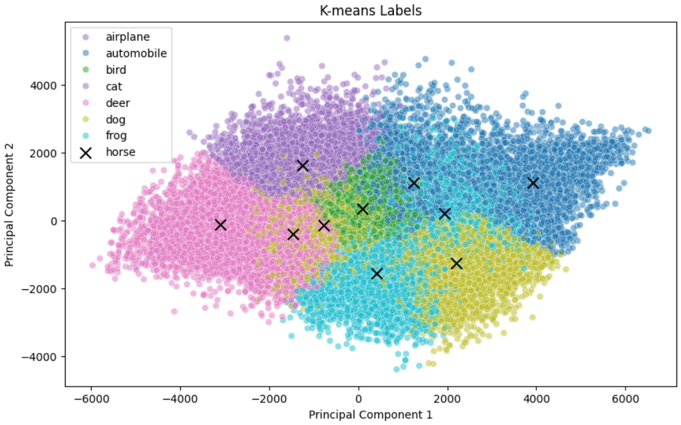


As we can see classes are mixed but not fully intermixed, which results in decent accuracy using k means, which is shown in the next plot





We can see the non-linearity in data, which could not be captured by k-means and thus results in poor accuracy



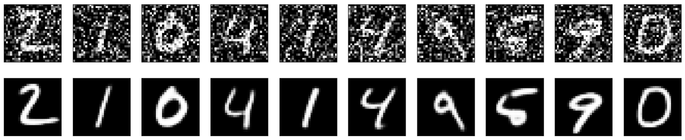
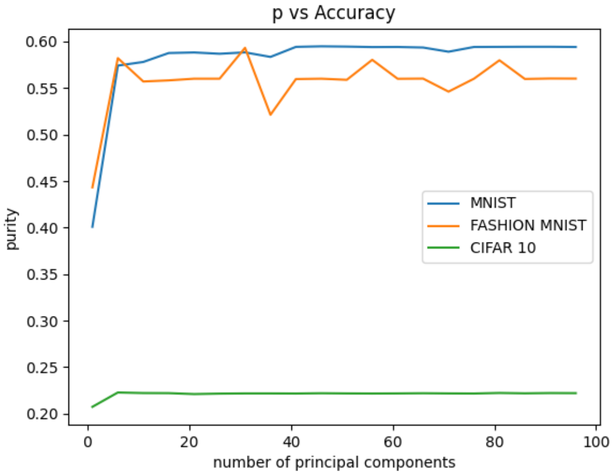


Fig. 1. Autoencoder result with Gaussian noise

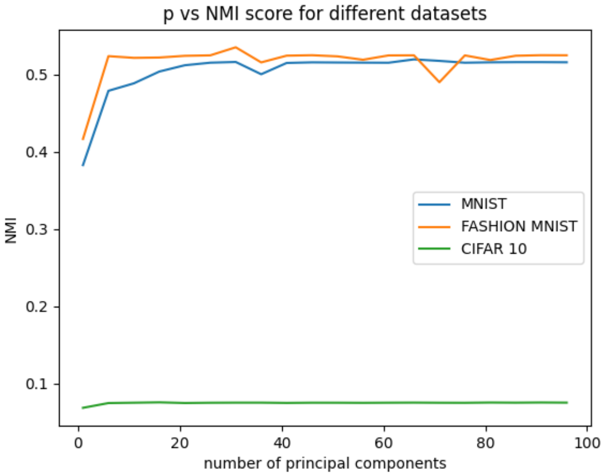
The above image shown is an image with Gaussian noise added to it. Below is the output of the autoencoder

B. CIFAR-10 Dataset

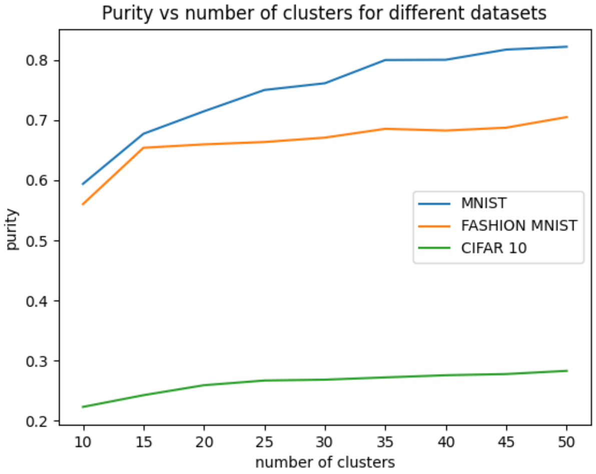
K-means alone performed poorly with only 21.9 percent accuracy, suggesting that CIFAR-10's higher complexity and colour information pose challenges for basic clustering algorithms. PCA combined with K-means showed no significant improvement, maintaining an accuracy of around 22 percent, which means clustering is solely done based on 1st feature of PCA.



Accuracy increases with an increase in the value of p but stagnates, which tells us that k-means focus on pixels with the most deviation



Graph is expected to be similar to purity vs p



VIII. Inferences

PCA may not effectively capture complex non-linear relationships within the data as a linear transformation, especially in datasets like MNIST and CIFAR-10. AutoEncoders, by contrast, have shown promise in capturing non-linear relationships, leading to improved clustering accuracy, especially in the MNIST dataset. Autoencoders trained on augmented images performed better than those which take the image as input and produce the image itself. Number of Clusters: Increasing the number of clusters in K-means improved accuracy. Further work could be done to merge those significant clusters into an original number of ground truth labels.

Possible reasons for better performance of hierarchical clustering to K-means. :

In K-means, the clustering depends on the initial seed placement

And K means is highly sensitive to outliers since, after each iteration, the recalculated centroids can be pushed toward the outliers.

Such a problem doesn't exist in hierarchical because the clustering is hierarchical with a bottom-up approach. We can also see that from the data plotted with original labels, clusters are not rounded but K-means clusters with the assumption that clusters are of spherical shape as it calculates the Euclidean distance from the centroids.

The Hierarchical algorithm is not grounded on any such assumption.

IX. Conclusion

A. *Effectiveness of Techniques*

AutoEncoders outperformed PCA in embedding high-dimensional data into lower-dimensional space, especially in capturing non-linear relationships. Combining embedding techniques like AutoEncoders with clustering algorithms like K-means or Agglomerative Clustering can improve clustering accuracy without labels.

B. *Dataset Complexity*

Dataset complexity plays a crucial role in the effectiveness of clustering algorithms. Simple datasets like MNIST yield higher accuracies than more complex datasets like CIFAR-10.

C. *Future Directions*

Work on autoencoder of CIFAR-10 and Fashion MNIST and find a way to merge a large number of clusters to the required number of clusters. Implementing better clustering algorithms after generating embedding space.

References

[1] SCAN: Learning to Classify Images without Labels https://arxiv.org/abs/2005.12320

[2] Deep Embedded K-Means Clustering https://arxiv.org/abs/2109.15149

[3] MNIST Autoencoder https://blog.keras.io/building-autoencoders-in-keras.html

[4] Fashion MNIST Autoencoder https://www.tensorflow.org/tutorials/generative/autoencoder

X. Individual Contributions

Both contributed equally to the project. Specifically:

Himanshu Singh:

- Read and understood relevant research papers.
- Implemented the codes.
- Interpreted the results of experiments and proposed methods for improving model performance.
- Experimented with different clustering techniques.

Jaleel Ahmed Radhu Khawaja:

- Read and understood relevant research papers.
- Tested the results and found solutions to enhance the model's robustness.
- Interpreted the results of experiments and proposed methods for better embedding.
- Experimented with different architectures of Autoencoders