PARTH VERMA – 2022CS11936 **AVIRAL SINGH – 2022CS11290**

Report on Representation Learning

Introduction

In this report, we present a VAE and a GMM model developed for digit classification. The model is trained on a dataset of unlabelled digit images and utilizes CNN to enhance performance.

Component 1: VAE

Model Architecture

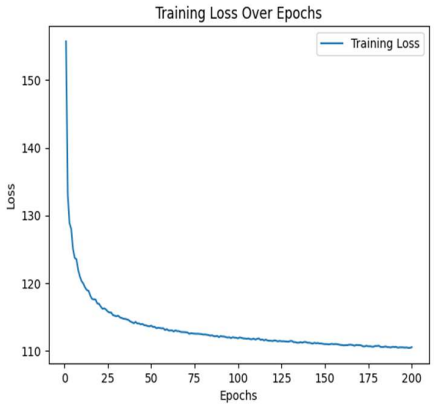
```
----------------------------------------------------------------
        Layer (type)               Output Shape         Param #
================================================================
            Conv2d-1          [-1, 64, 14, 14]           1,088
              ReLU-2          [-1, 64, 14, 14]               0
            Conv2d-3           [-1, 128, 7, 7]         131,200
       BatchNorm2d-4           [-1, 128, 7, 7]             256
              ReLU-5           [-1, 128, 7, 7]               0
            Linear-6                  [-1, 20]         125,460
            Linear-7                  [-1, 20]         125,460
            Linear-8                [-1, 6272]         131,712
   ConvTranspose2d-9          [-1, 64, 14, 14]         131,136
  ConvTranspose2d-10           [-1, 1, 28, 28]           1,025
================================================================
Total params: 647,337
Trainable params: 647,337
Non-trainable params: 0
----------------------------------------------------------------
```

Reconstructed Images on given validation dataset

Average MSE: 0.0192
Average SSIM: 0.7212

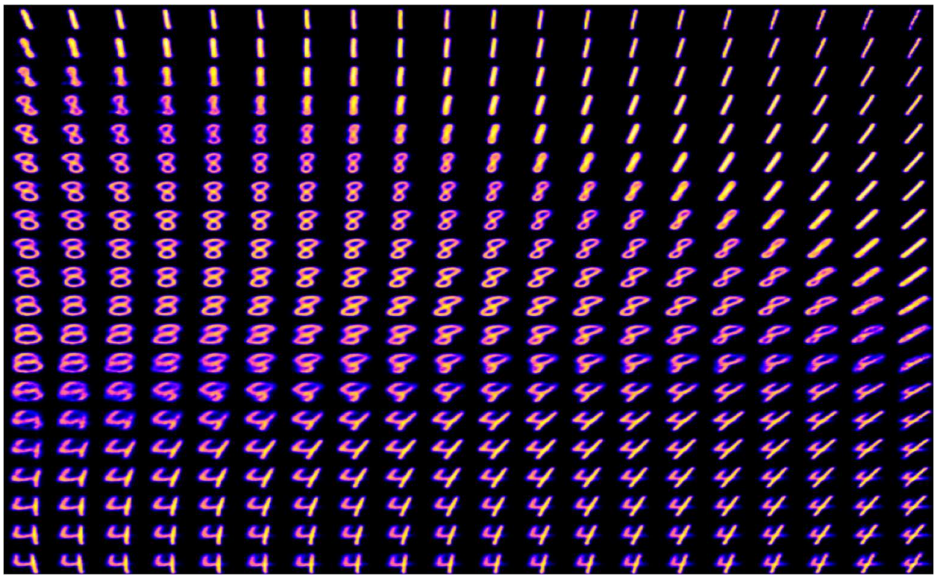
Using CNN improved the performance of VAE model, as compared to MLP architectures.

Training Loss vs Epoch



Generating new Images

The following are the generated images which shows VAE generative performance.

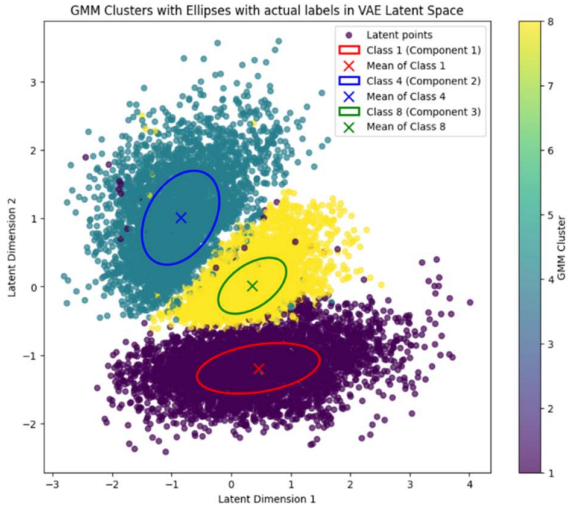


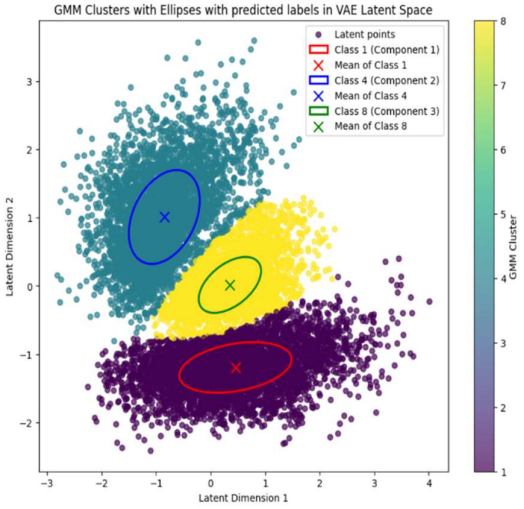
Component2: GMM

This project implements a Gaussian Mixture Model (GMM) on the latent space of a Variational Autoencoder (VAE) to cluster images into classes. The GMM, initialized with class-specific latent means from validation data, iteratively refines component means, covariances, and weights using the Expectation-Maximization (EM) algorithm. The given training dataset was spitted into 80 – 20 ratios to create a large validation set for tuning hyperparameters.

Clusters were initialized using the given 15 labelled images, and the GMM was trained using the spitted training set. Later its performance was evaluated by classifying images of the test set

GMM Clusters





Analysis of Clusters

- **Cluster Separation:** The VAE-GMM model effectively separates the data into distinct clusters, indicating that the VAE has learned a meaningful latent representation suitable for clustering. They also show expected normal distribution
- **Overlap and Misclassification:** Some overlap is noticeable, particularly between the clusters for "Class 4" and "Class 8." This overlap suggests potential areas for misclassification, as data points near these boundaries may be ambiguously assigned.
- **Cluster Compactness:** The cluster for "Class 1" is relatively compact, indicating a tighter grouping of latent points and potentially more consistent classification. In contrast, "Class 8" is more spread out, suggesting higher variability within this class.
- **Model Performance Summary:** The VAE-GMM combination generally achieves clear clustering. However, reducing overlaps and refining the latent space could improve classification accuracy, particularly for classes with more variability.