Assignment 3, part 2
Report

Disha (2022CS11118) and Avilasha Mandal (2022CS11631)

VAE Model Architecture

The VAE consists of two main components: an **encoder** and a **decoder**. The encoder maps input images into a latent space, while the decoder reconstructs images from the latent representations. The latent space is modeled as a distribution (Gaussian), which helps in generating new data by sampling latent variables.

Encoder:

- The input image is flattened into a 1D vector and passed through a fully connected (FC) layer with ReLU activation.
- The encoder produces two outputs: **mean (μ)** and **log-variance (log(σ2))**, which represent the parameters of the Gaussian distribution in the latent space.

Reparameterization Trick:

- To allow backpropagation, the latent variable z is sampled using the reparameterization trick: $z=\mu+\sigma\cdot\epsilon$ where $\epsilon\sim N(0,I)$ is a noise term.

Decoder:

- The latent variable z is passed through a decoder, which consists of a fully connected layer followed by a sigmoid activation function to produce the reconstructed image.

VAE Loss Function

1. Reconstruction Loss (Binary Cross-Entropy): Measures how well the VAE can reconstruct the input image:

$$\mathcal{L}_{\text{reconstruction}} = -\sum_i x_i \log(\hat{x}_i) + (1 - x_i)\log(1 - \hat{x}_i)$$

where xi is the original image pixel, and x^i is the reconstructed pixel.

2. KL Divergence Loss: This regularizes the latent space by minimizing the difference between the learned distribution and a standard normal distribution:

$$\mathcal{L}_{\text{KL}} = -\frac{1}{2} \sum_{j} \left(1 + \log(\sigma_j^2) - \mu_j^2 - \sigma_j^2\right)$$

where μ_j and σ_j^2 are the mean and variance of the learned latent distribution.

The total VAE loss is the sum of these two components:

$$\mathcal{L}_{\text{VAE}} = \mathcal{L}_{\text{reconstruction}} + \mathcal{L}_{\text{KL}}$$

Results and Observations

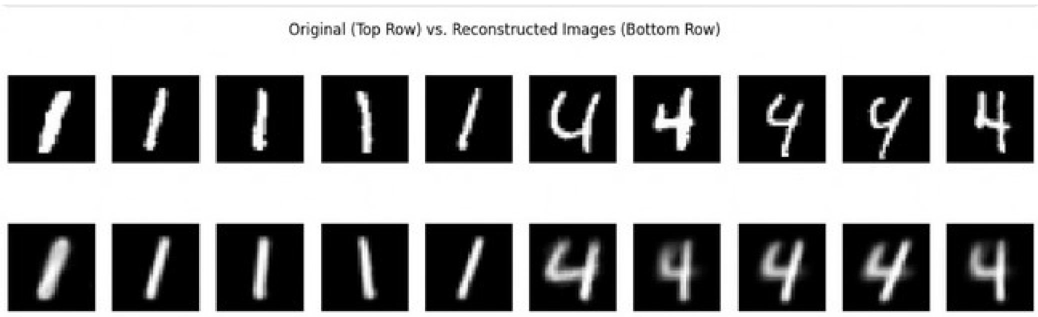
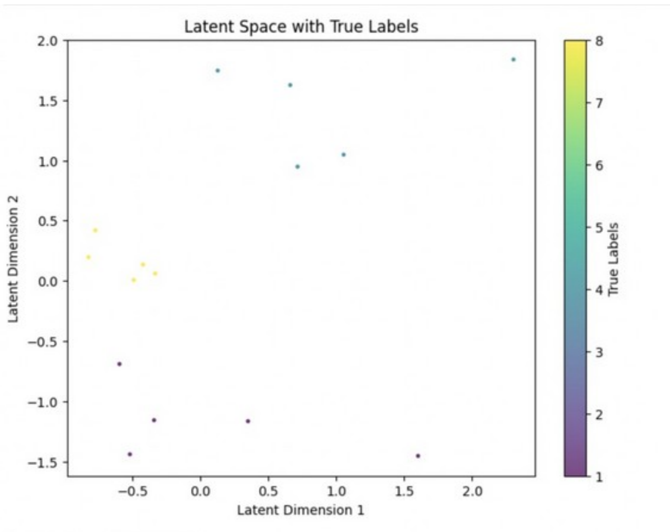


Image Reconstructions:
Below are examples of reconstructed images from the validation dataset. The VAE demonstrates strong reconstruction performance, with minimal differences between the original and reconstructed images.

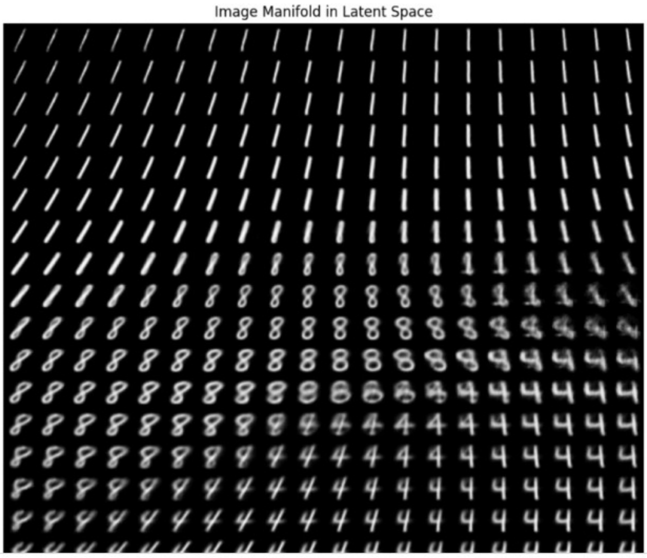
Latent Space Scatter Plot:
The latent space of the VAE is visualized in a 2D scatter plot, where each point corresponds to the latent vector of an image. This visualization shows how the VAE has structured the data in the latent space and whether it has clustered similar images together.

The scatter plot shows clusters of similar labels are loosely formed, but there is overlap, indicating that the latent representation does not perfectly separate classes.



Generated Images:

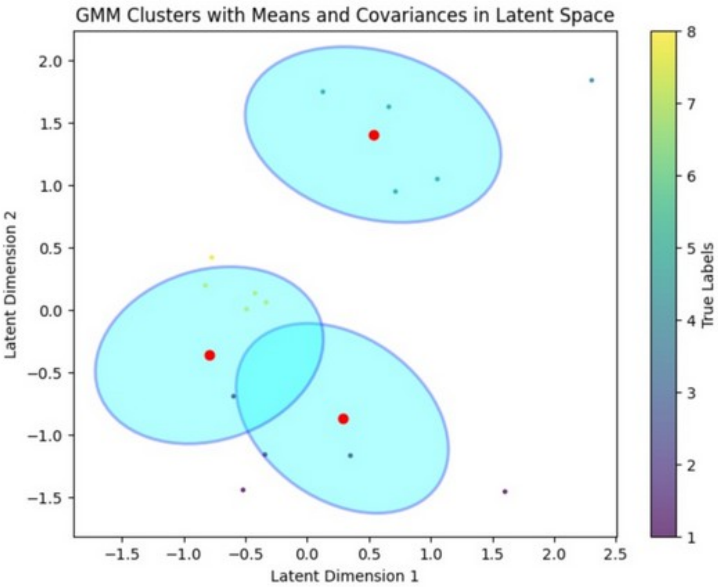
The following images were generated by sampling from the latent space and passing the sampled latent vectors through the decoder. These images demonstrate the generative capability of the VAE.



This grid visualizes digit images in the latent space, with smooth transitions between digits showing continuity. Different digit types cluster in distinct regions, indicating that the model captures underlying patterns, though some overlap causes digit blending.

GMM Classification Results:

The VAE's latent vectors were used for classification using a Gaussian Mixture Model. Below are the performance metrics (accuracy, precision, recall, F1 score) for the GMM-based classification.



The GMM clustering in the latent space shows partial alignment with the true labels, as represented by the color-coded data points. Each light blue ellipse indicates a cluster's spread and orientation, with red dots marking the cluster centroids. Some overlap between clusters and labels suggests that the latent space does not fully separate the classes, meaning the GMM captures broad groupings but may struggle with precise distinctions.

Conclusion

The VAE successfully learns to encode images into a low-dimensional latent space and reconstruct them with high fidelity. The latent space visualization confirms that the VAE has captured meaningful structure in the data, while the generated images demonstrate its ability to create new samples. The application of GMM on the latent vectors also shows that the VAE's learned representations are useful for downstream tasks such as classification.