Latent Space Exploration Using a Variational Autoencoder (VAE)

TUTORIAL 4

SUBMITTED BY

VARUN GADI

RA2211027010203

AD2

## DSBS

Variational Autoencoders (VAEs) are a class of deep generative models that learn to encode high-dimensional data into a lower-dimensional latent space. Unlike traditional autoencoders, VAEs incorporate a probabilistic approach to learning the latent distribution, allowing for more controlled generation of new data.

In this tutorial, we explored the latent space of a trained VAE model on the Fashion MNIST dataset. The goal was to:

1. Visualize the structure of the latent space.

2. Generate new samples by sampling from the latent space.

3. Manipulate the latent space to analyze how the learned features correspond to different attributes in images.

| Input | Encoder | Latent Space | Decoder | Output |
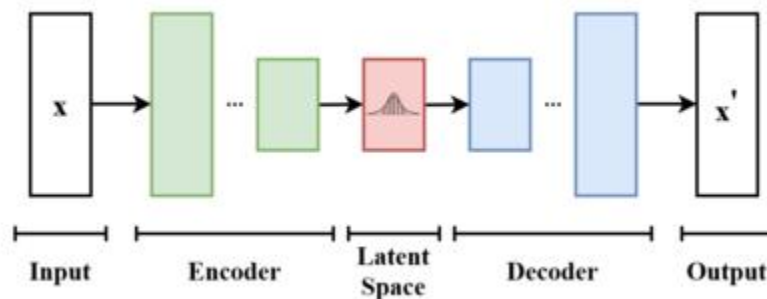
Implementation Details

Dataset and Preprocessing

We used the Fashion MNIST dataset, which consists of 60,000 training and 10,000 test grayscale images of clothing items categorized into 10 classes:

- T-shirt, Trouser, Pullover, Dress, Coat, Sandal, Shirt, Sneaker, Bag, and Ankle Boot.

Each image is 28×28 pixels, and we applied normalization to scale pixel values between -1 and 1 to enhance stability during training.

Fashion MNIST Samples
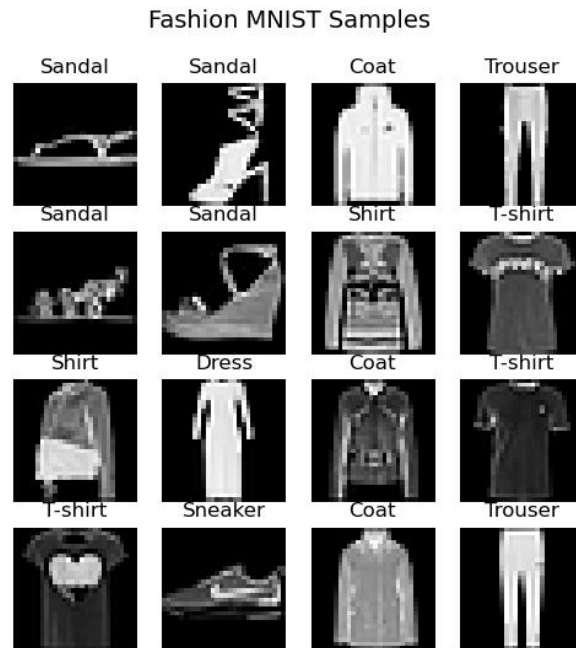

Variational Autoencoder (VAE) Architecture

The VAE consists of:

- Encoder: Compresses the input into a lower-dimensional latent space.

- Reparameterization Trick: Introduces stochasticity by generating a latent representation from a mean and variance.

- Decoder: Reconstructs the input image from the latent representation.

Loss Function

The VAE loss function combines:

1. Reconstruction Loss: Ensures the decoded output is similar to the original input.

2. KL Divergence Loss: Regularizes the latent space by enforcing a normal distribution.

VAE Training Loss
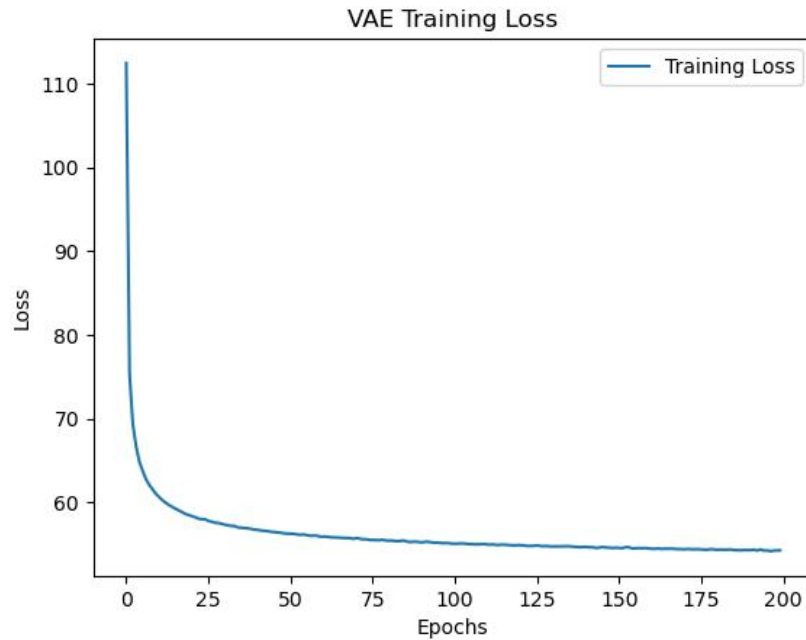
Results & Analysis

1. Interpolation in Latent Space

One way to analyze the latent space is to interpolate between two points in the space and visualize the gradual transformation. We took two different clothing items, mapped them to their respective latent representations, and generated intermediate images by smoothly transitioning between them. This helped us understand how the model learns the variations between different classes.

Original (Top) vs Reconstructed (Bottom) Images
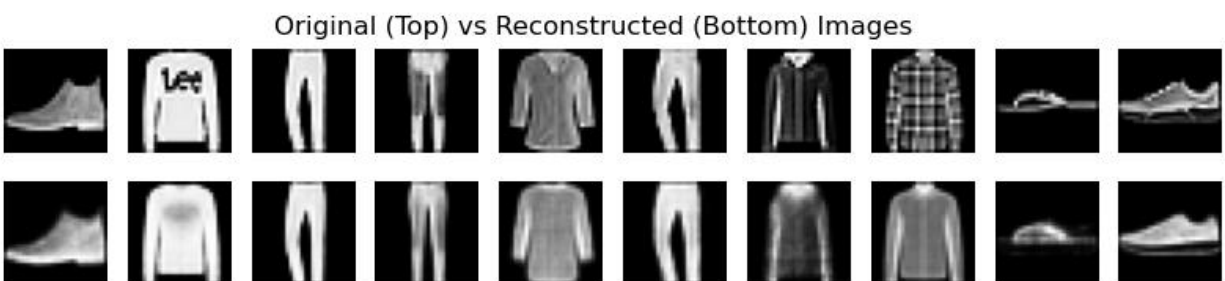
2. Latent Space Visualization

To see how well the VAE clusters different clothing types in the latent space, we used:

- t-SNE (t-Distributed Stochastic Neighbor Embedding): A dimensionality reduction technique that maps high-dimensional latent vectors to a 2D plot. This helped in observing how different classes of images were organized in the latent space.

- PCA (Principal Component Analysis): Another method to reduce dimensionality and visualize how the latent variables separate the dataset.

The results showed that certain categories (e.g., sneakers and sandals) had distinct clusters, while others (e.g., shirts and coats) overlapped slightly, indicating similarities in their visual features.

Latent Space Representation using t-SNE
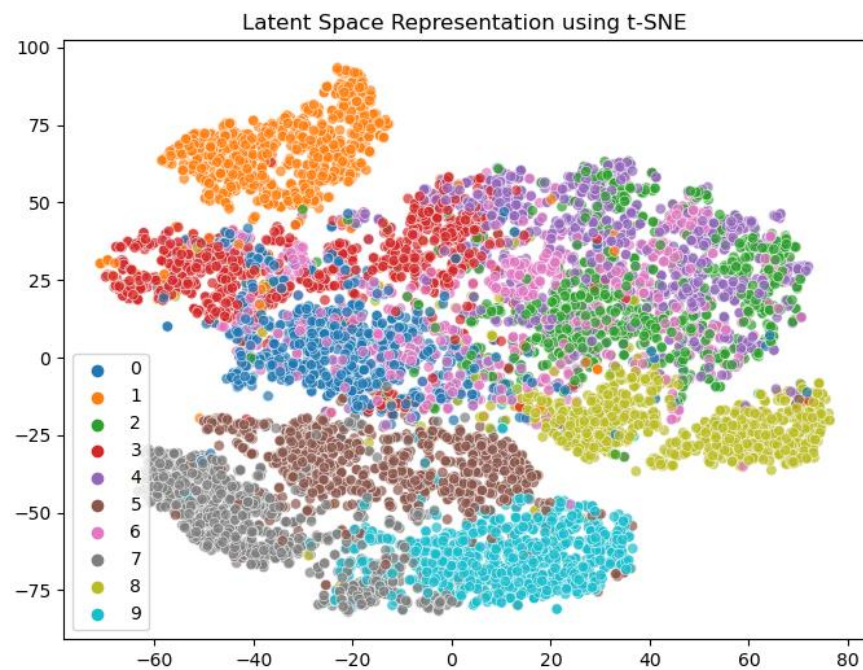
3. Generating New Samples

To assess how well the VAE generalizes, we sampled random points from the learned latent space and passed them through the decoder. The generated images resembled real clothing items, showing that the model successfully captured meaningful representations of Fashion MNIST classes.

Generated Images from Latent Space

Conclusion

This exploration of the VAE's latent space provided insights into how deep generative models encode information and how we can manipulate the latent space to control image generation.

Key Takeaways:

- The VAE successfully learned a structured latent space, allowing smooth interpolations between different clothing items.

- Latent space visualization (t-SNE, PCA) helped reveal how the model organizes different classes.

- Generating new images from sampled latent vectors showed realistic outputs, confirming that the VAE captures meaningful representations.

- Some classes had overlapping clusters, indicating that the model could be improved by adjusting the latent space dimension or fine-tuning hyperparameters.

Future Enhancements:

- Experimenting with different latent space sizes to improve the separation of classes.

- Using higher-resolution datasets to generate more detailed images.

- Exploring Conditional VAEs (CVAE) to generate specific clothing items based on class labels