Assignment 3 - Variational Autoencoders (VAEs)

Narahara Chari Dingari, Ph.D.

DS552 - Generative AI

1. Theory Questions:

- **Q1: Why is the KL Divergence term important in the VAE loss function?**
- **Q2: How does the reparameterization trick enable backpropagation through the stochastic layers of a VAE?**
- **Q3: Why does a VAE use a probabilistic latent space instead of a fixed latent space?**
- **Q4: What role does KL Divergence play in ensuring a smooth latent space?**

2. Coding Tasks:

Task 1: Modify the VAE architecture to use convolutional layers for both the encoder and decoder, and train it on the CIFAR-10 dataset. This modification will allow the model to capture spatial relationships within images more effectively, improving its ability to generate high-quality images. After training, compare the generated images with those from a fully connected VAE.

Task 2: Using the trained VAE, interpolate between two images in the latent space and generate intermediate images. This demonstrates how smoothly the model can transition between different data points. Visualize and display the results, showing the interpolated images in a grid format to observe the transformation.

Task 3: Train the VAE on a new dataset of your choice (e.g., CelebA for faces), and visualize generated samples. Experiment with sampling from different regions of the latent space and analyze how the generated outputs vary based on different latent vectors.

Coding Task Explanation: In this assignment, the VAE will be expanded by incorporating convolutional layers in both the encoder and decoder networks. Convolutional layers are particularly effective for image data, as they can capture spatial hierarchies better than fully connected layers. By using this architecture on the CIFAR-10 dataset (a dataset of small, colorful images), the VAE will be able to learn a more effective latent representation, which should improve the quality of generated images.

The second task involves interpolation in the latent space. By interpolating between two points in the latent space (corresponding to two different images), we can observe how smoothly the VAE transitions between two images. This task highlights the structure of the learned latent space and shows the generative capabilities of the VAE.

Finally, applying the VAE to a new dataset (such as CelebA, a dataset of celebrity faces) and visualizing the generated images allows us to see how well the model generalizes to different kinds of data and how the latent space is structured for more complex, real-world images.