Christopher Santorelli
DS552 – Homework 3 Theory Questions

1. Theory Questions:

Q1: Why is the KL Divergence term important in the VAE loss function?

The Kullback-Leibler (KL) Divergence term is important to the VAE loss function because it is responsible for minimizing the difference between the learned latent distribution which is defined by the encoder and a standard Gaussian Distribution. By minimizing the difference, it keeps the latent space smooth and continuous which ensures that the data generated is meaningful from any point within the latent space.

If the KL divergence term was removed, the learned representations would become disjointed or overfitted, reducing the generative capabilities of the VAE leading to poor generalization and degraded generative quality.

Q2: How does the reparameterization trick enable backpropagation through the stochastic layers of a VAE?

The reparameterization trick enables backpropagation through the stochastic layers of a VAE by producing the deterministic parameters, mean (μ) and the standard deviation (σ) which are functions of the input and the network parameters as an alternative to sampling the latent variable (z) straight from the probability distribution.

The trick then introduces "controlled" randomness via the noise term (ϵ) which is drawn from a standard normal distribution N (0,1). The full equation for reparameterization ($z = \mu + \sigma * \epsilon$) separates the randomness contained in ϵ from the mean plus the standard deviation, because the mean and standard deviation are differentiable, noise (ϵ) is independent of those parameters, allowing the full equation to be differentiable. This allows gradients to flow through the mean and standard deviation through backpropagation, enabling the network to learn even though there is stochastic sampling.

Q3: Why does a VAE use a probabilistic latent space instead of a fixed latent space?

The VAE uses a probabilistic latent space instead of a fixed latent space such that it can generate diverse and varied outputs from the same input. The encoder maps to a distribution instead of to a fixed point allowing the decoder to sample different points from its continuous latent space.

Q4: What role does KL Divergence play in ensuring a smooth latent space?

The Kullback-Leibler (KL) Divergence ensures a smooth latent space by regularizing the learned latent distributions towards a standard Gaussian distribution.