

1. Why is the KL Divergence term important in the VAE loss function?

KL Divergence regularizes latent space by minimizing the difference between the learned distribution and the standard Gaussian. This forces the distribution to stay close to a known prior, and without it, the latent space can become messy or discontinuous.

2. How does the reparameterization trick enable backpropagation through the stochastic layers of a VAE?

Because sampling from the latent space disrupts backpropagation, reparameterization rewrites the sampling step using a Gaussian distribution allowing gradients to flow through the distribution while keeping a fixed noised term.

3. Why does a VAE use a probabilistic latent space instead of a fixed latent space?

By using a probabilistic latent space the data is encoded as a distribution which enables the generation of diverse outputs from the same input by sampling different areas of the distribution, rather than only being able to produce identical data from traditional latent space.

4. What role does KL Divergence play in ensuring a smooth latent space?

KL Divergence pulls the encodings toward a shared prior which prevents different inputs from being encoded into extremely different regions and encourages the distributions to overlap in a common latent space. KL Divergence also fills in the gaps between data points by staying near the prior mean.