VARIATIONAL AUTOENCODERS (VAE's) NOTES

Variational Autoencoders (VAEs) are a type of generative model that can generate new data by learning a compressed representation of the input data and then decoding it back to the original data space. Here's how VAEs work, explained in simple terms:

Key Components of a VAE

1. **Encoder**:
    - The Encoder is a neural network that takes the input data (like an image) and maps it to a compressed representation called the **latent space**. This latent space is typically represented by a mean vector and a variance vector, which together define a **probability distribution** over the latent space.
2. **Latent Space**:
    - The latent space is a lower-dimensional space where the VAE represents the data. Instead of mapping inputs to a fixed point in this space, VAEs map inputs to a distribution (usually a Gaussian distribution).
    - The latent variables (latent codes) are sampled from this distribution, which introduces some randomness. This sampling process is crucial for the generative capability of VAEs.
3. **Decoder**:
    - The Decoder is another neural network that takes a point from the latent space and maps it back to the original data space. It reconstructs the data, aiming to make it as close as possible to the original input.

Generating New Data with VAEs

1. Learning the Distribution:

- During training, the VAE learns to encode data into the latent space and decode it back accurately. The loss function used in training has two main components:
    - **Reconstruction Loss**: Measures how well the output data matches the input data.
    - **Kullback-Leibler Divergence (KL Divergence)**: Measures how close the learned latent space distribution is to a standard normal distribution (a Gaussian distribution with mean 0 and variance 1).

2. Sampling from the Latent Space:

- After training, the VAE has learned a meaningful latent space that captures the underlying structure of the data. To generate new data, we can:
    - Sample a random point (latent vector) from the standard normal distribution (which the latent space has been regularized to resemble).
    - Feed this latent vector into the Decoder network.

3. Decoding the Sampled Latent Vector:

- The Decoder network then transforms the sampled latent vector into new data, such as an image, sound, or text, depending on the training data used.

How Does It Generate New Things?

The key to VAEs generating new things lies in the **latent space**. By sampling from the latent space and decoding the samples, VAEs can create new, previously unseen data. Because the latent space is continuous and the VAE learns a smooth mapping from this space to the data space, slight changes in the sampled latent vectors can result in meaningful variations in the generated outputs.

Summary

- **VAEs encode data into a probabilistic latent space** using an Encoder, **sample from this space**, and then **decode these samples back into data** using a Decoder.
- They generate new data by sampling new latent vectors from the learned latent distribution and decoding them.
- This process allows VAEs to create novel variations that were not present in the training data but are similar in style and content.

VAEs are powerful because they not only compress data into a latent representation but also provide a structured way to explore and generate new data samples.