Deep Learning Lab 8

Jack Hilton-Jones (jhj1g23)

I. Exploring the latent space of a VAE

Using a Variational Autoencoder (VAE) an image is created by systematic sampling, the VAE is trained on the Fashion MNIST. With this VAE model, the decoder network generates images by sampling from $-4\sigma to +4\sigma$ in both dimensions. For each dimension, 21 samples are created, and each reconstruction is drawn into a 528×528 image according to the z vector that generated it. The image is seen in Figure 1.

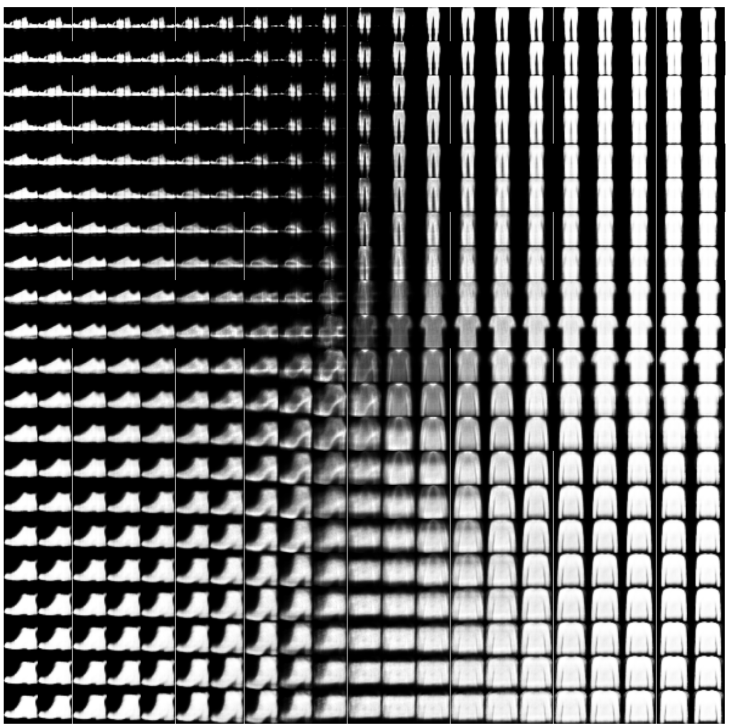

Fig. 1. VAE with systematic sampling from the Fashion MNIST

II. Exploring the code space of a standard controller auto-encoder

In this section, a standard autoencoder is used to generate images. This uses the decoder part of the network independently. The latent space is 2 dimensional and an image is created with the same number of samples and range per dimension. The image is seen in Figure 2

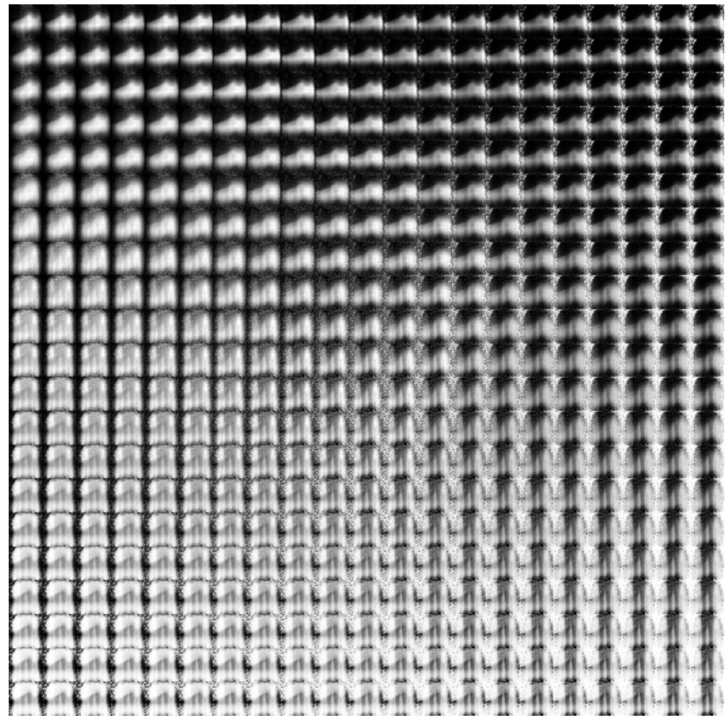

Fig. 2. Standard auto-encoder with systematic sampling from the Fashion MNIST

Observing the two images produced by the VAE and the standard auto-encoder there are distinct differences. Firstly the diversity of the images is evident, the number of different images generated by the VAE surpasses the standard auto-encoder. This is due to the VAE modelling the latent space as a continuous and probabilistic distribution. Sampling from this distribution uses the reparameterisation trick which enables gradient descent. This allows the VAE to approximate the true data distribution and the ability to switch smoothly between points in the latent space to produce varied output while capturing features in the dataset. The VAE can vary the latent variables to generate new samples. In comparison, the simple autoencoder is not as diverse, the model cannot learn the distribution over the latent variables but instead, the model directly maps from the input space to a compressed representation. Therefore the generative abilities of the autoencoder produce less variation and more blurring in the outputs.

The simple autoencoder produces poorer performance compared to the VAE, as many of the images are ambiguous and cannot be deciphered. This could be due to the simple autoencoder learning a latent space that minimises the reconstruction error leading to lower performance when generating new data. The VAE is more robust as it includes a regularisation term. The Kullback-Leibler (KL) divergence is a measure of how much information is lost using Q to represent P, this ensures the latent distribution approximates a standard normal distribution. The model is then prevented from lying far from the predefined distribution when structuring the latent distribution, allowing the model to generate new data points effectively. It also reduces overfitting to the training data, as the model produces a general representation of the data that captures the properties of the data set. The simple autoencoder produces worse interpolation switches between images as the outputs are sometimes unrealistic, due to the structure of the latent space not following a normal distribution. In conclusion, the poorer performance of the simple autoencoder is due to the direct mapping of the inputs to a 2-dimensional latent space, not containing a regularisation term and lacking generative properties that the VAE contains that define the probability distribution.