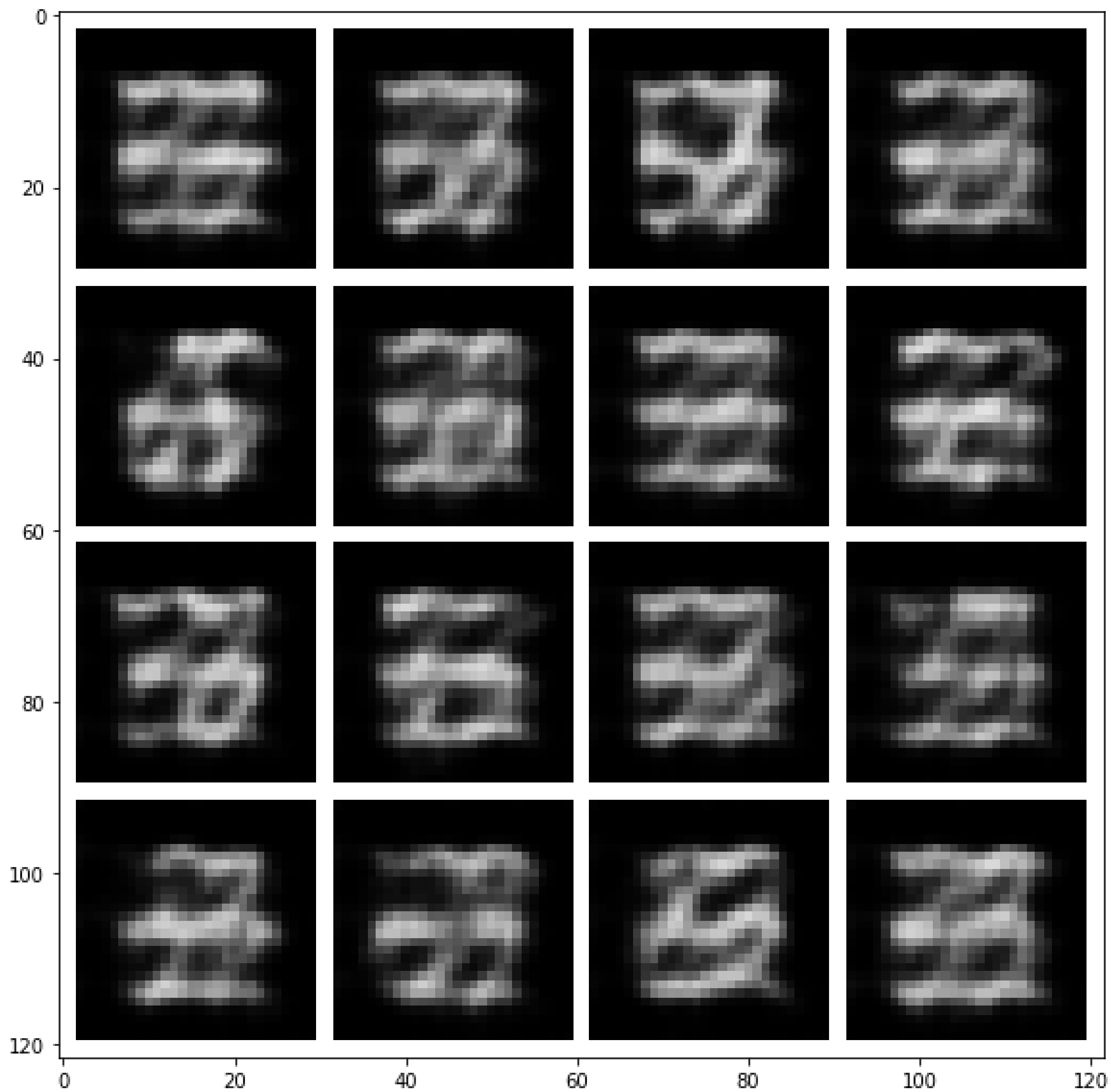
HW2 AE sample images -

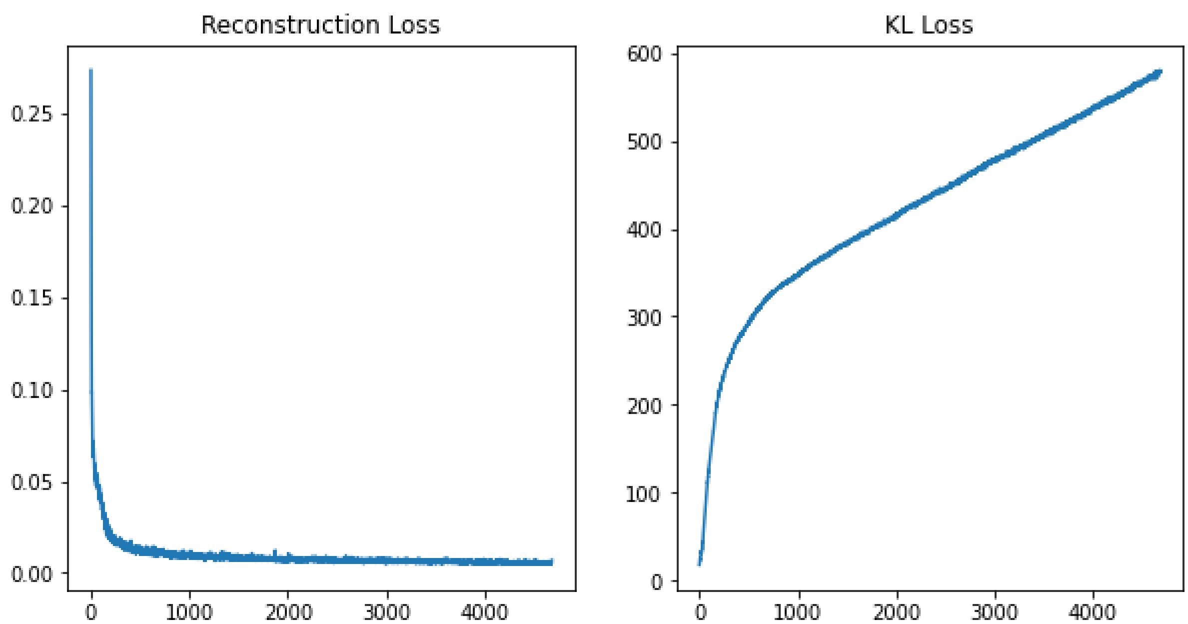


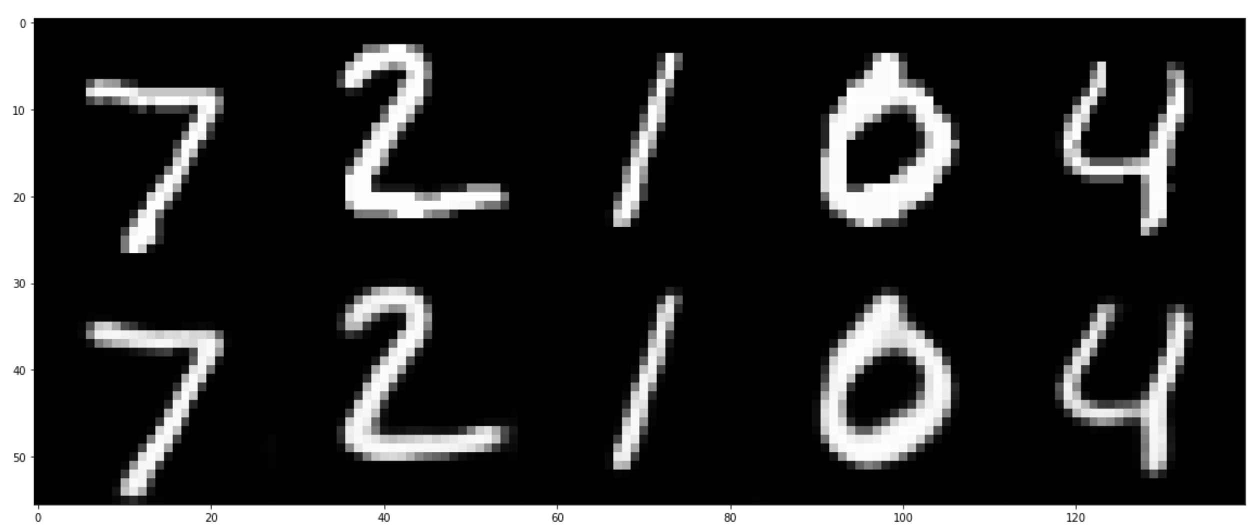
Answer to AE sampling inline question -

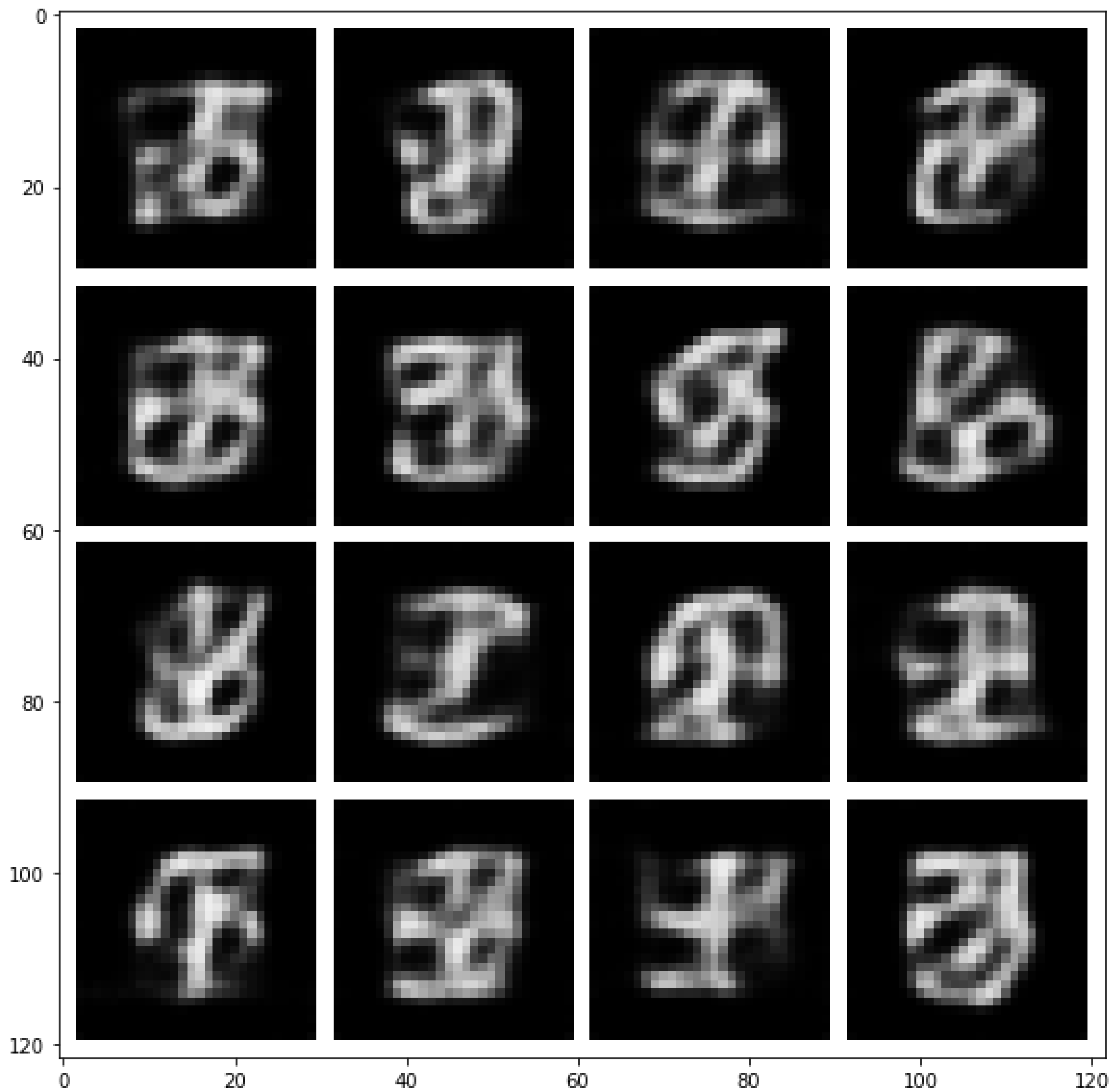
The issue with autoencoders being good generative models is that the encoded latent space may not allow easy interpolation due to a high amount of irregularity/discontinuity. The latent space is not regular, it depends on the distribution of data, model architecture and the dimension of latent space. The model does not produce organized latent space as it does not learn the distribution of the input data during the training process which aids to generate good samples. The autoencoder can perform well only if the sample embedding lies within the encoded space. It can perform well for reconstruction but not for generating images from new sample embeddings. Thus, the results obtained are very random and not well-defined digits.

HW2 VAE images -

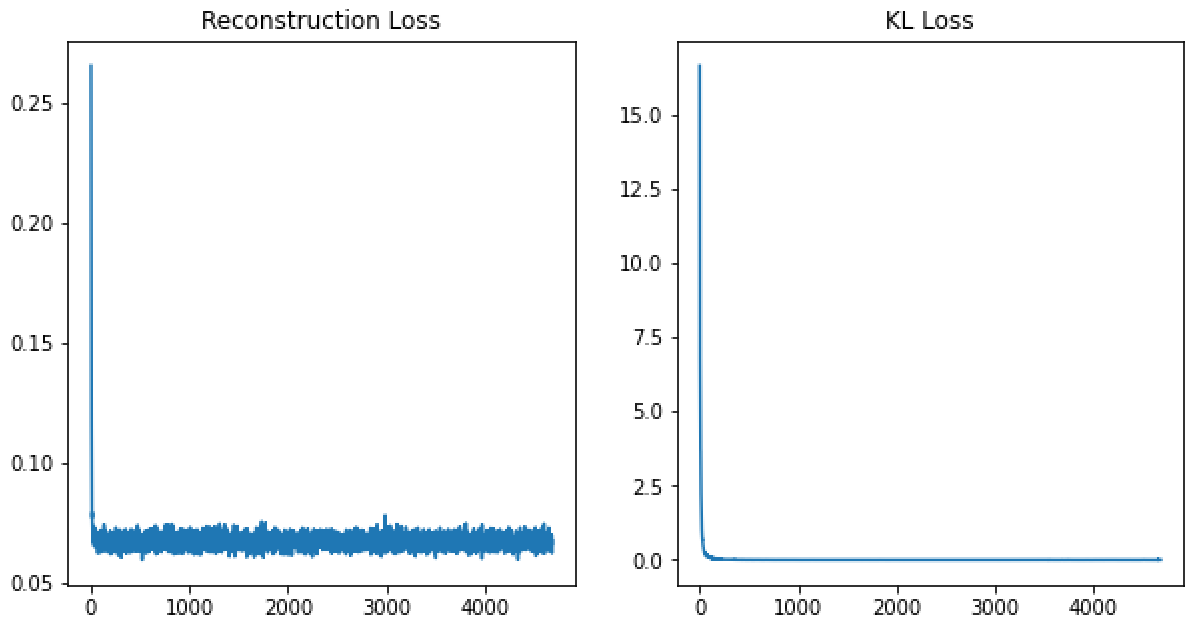
1) For Beta = 0:

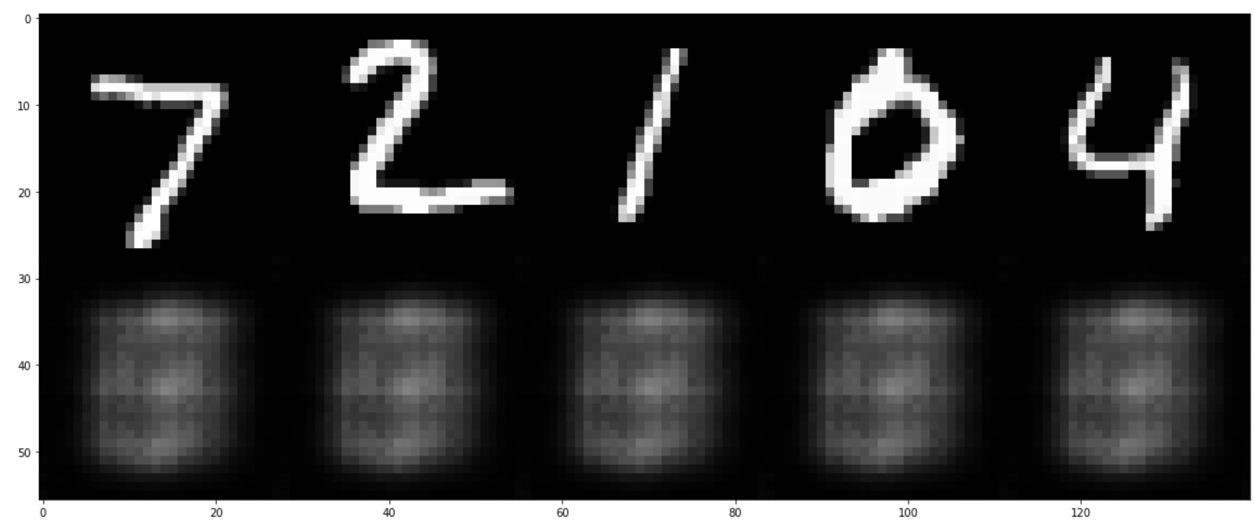


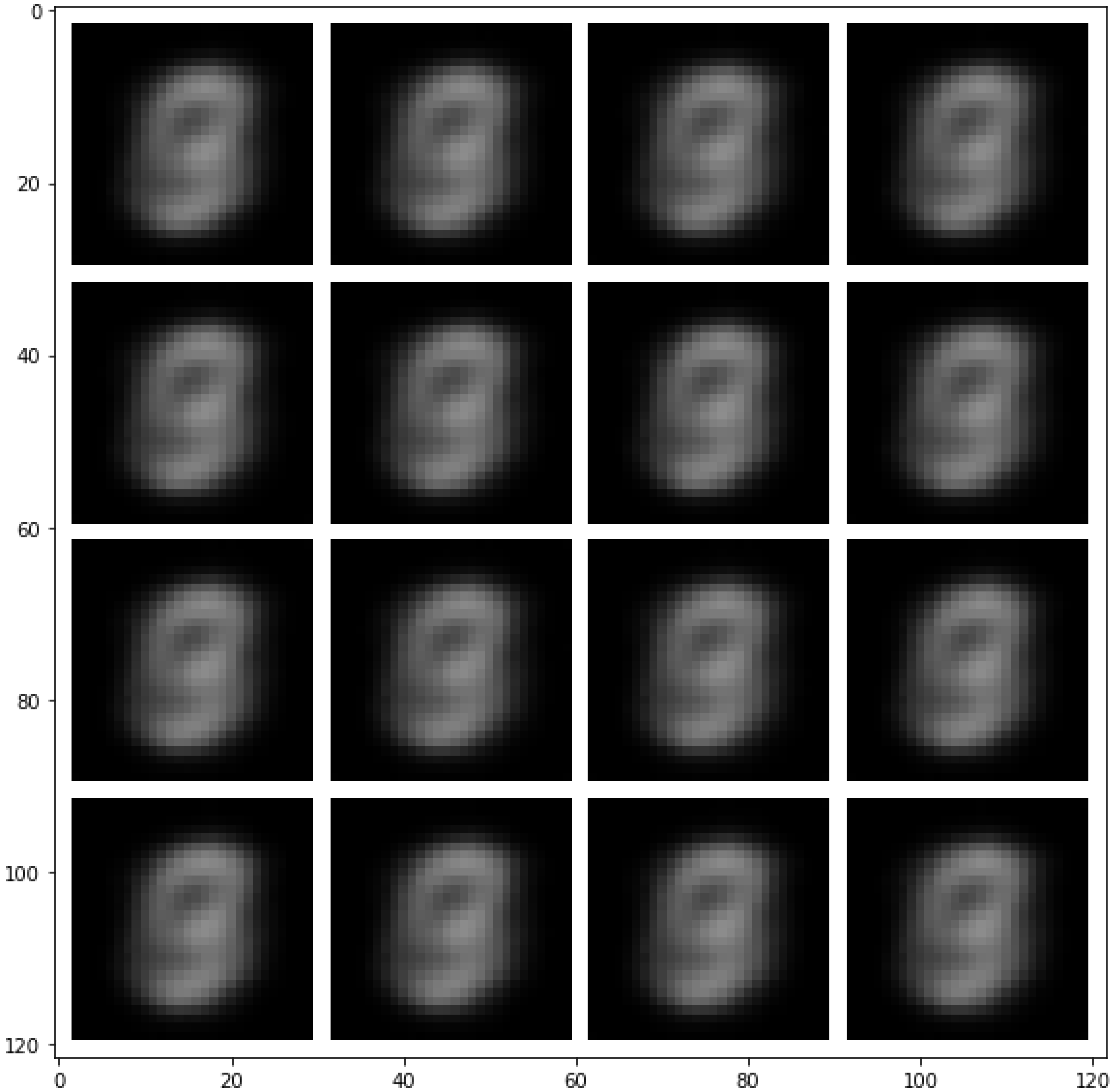




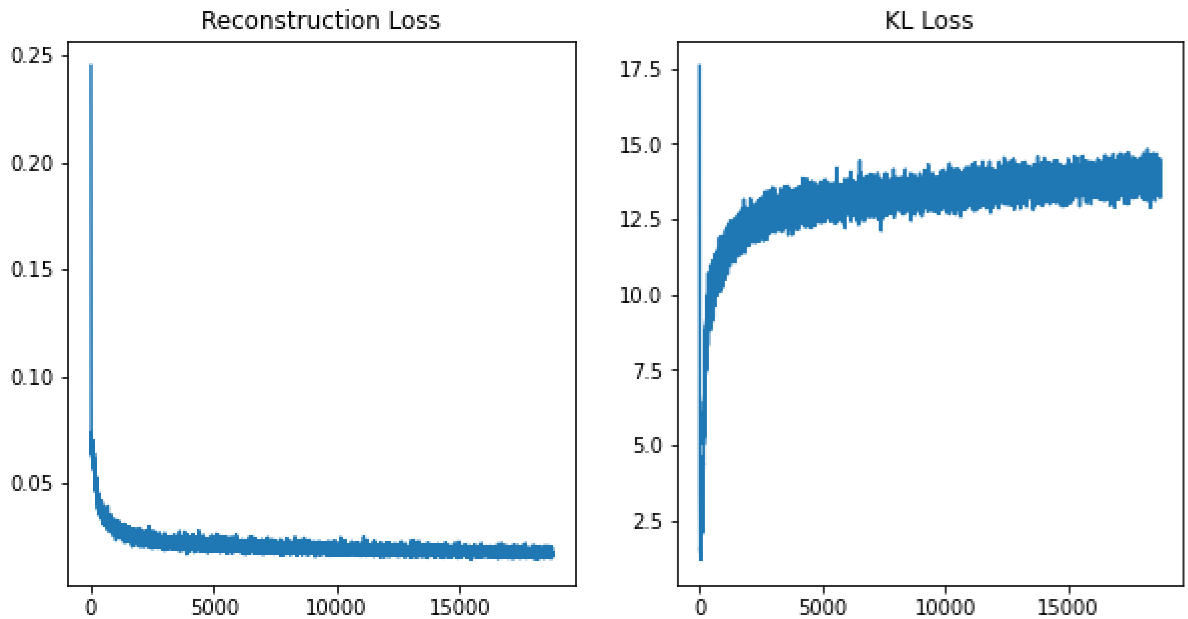
2) For Beta = 10:

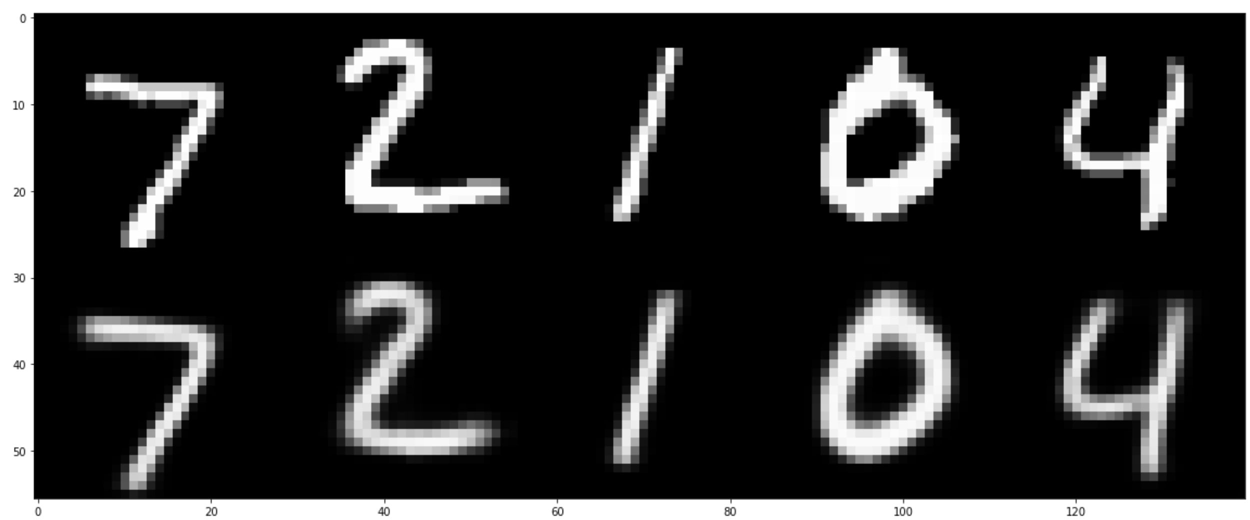

Reconstruction Loss
0.25
0.20
0.15
0.10
0.05
0
1000
2000
3000
4000
KL Loss
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0
1000
2000
3000
4000

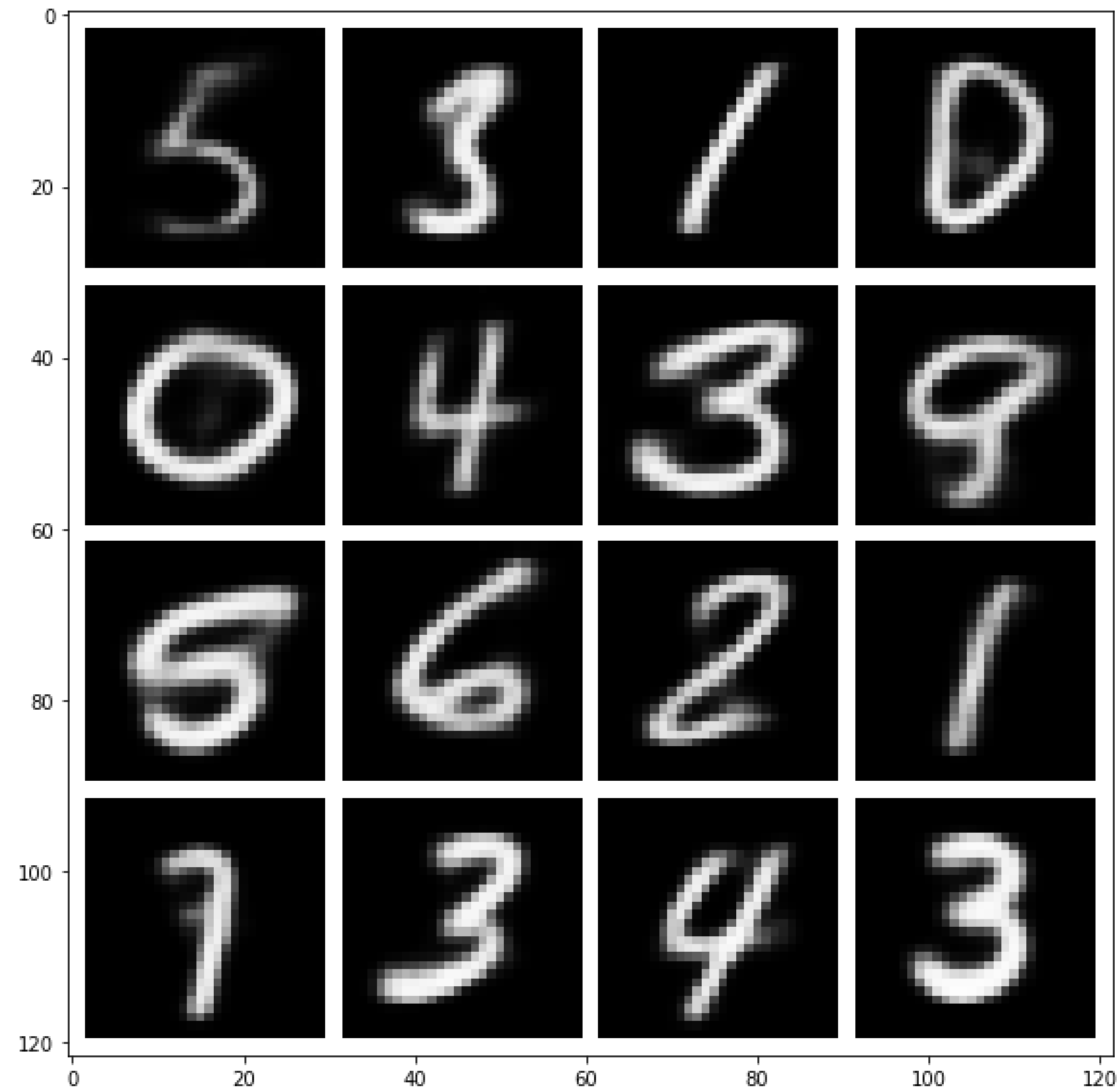

0
10
20
30
40
50
0
20
40
60
80
100
120



3) For Beta = 0.001


Reconstruction Loss
0.25
0.20
0.15
0.10
0.05
0
5000
10000
15000
KL Loss
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0
5000
10000
15000


0
10
20
30
40
50
0
20
40
60
80
100
120



Answers to the inline questions for Beta values 0 and 10-

1) For Beta = 0:

Due to the irregularity of the latent space in autoencoders, points sampled from such a latent space generate different decoded outputs. Variational Autoencoder adds a regularization term using KL Divergence which pushes the output of the encoder to be as close as possible to the prior distribution. By setting Beta=0, VAE almost behaves like autoencoders without the regularization term beta on the latent space distribution. Consequently, it does not have a good generative process, any sample point after decoding might not produce clear outputs. We can observe that the generated samples in this case are not good and the KL loss keeps increasing. The reconstruction image is the same as that from the autoencoder, because by setting the beta term to zero, we are still retaining the reconstruction loss term and thereby leads to the same objective function as that for Autoencoders.

2) For Beta = 10:

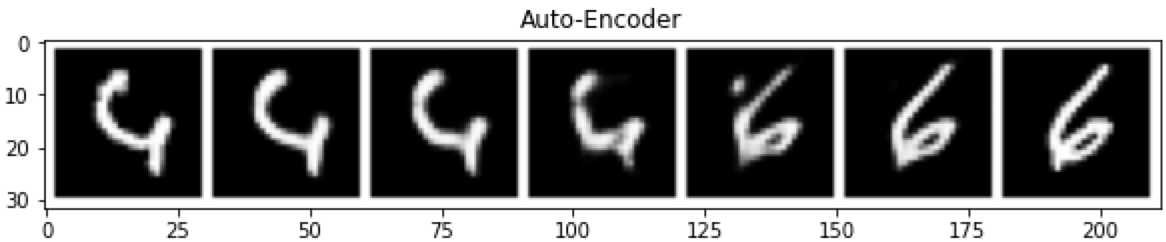
We can observe that the reconstructions and the samples produced are very blurred. Any Beta value higher than 0 applies a stronger constraint to the latent distribution, thus it can generate efficient latent encoding and the representations are disentangled. But if the Beta value is too high like 10, there will be a trade off between reconstruction quality and the disentanglement. We can clearly see that increasing the beta value to 10 imposes a stronger constraint on the latent distribution at the cost of the reconstruction quality(blurriness level). By constraining the latent distribution with high beta value and maintaining a low KL Loss, it decreases the reconstruction ability of the VAE.

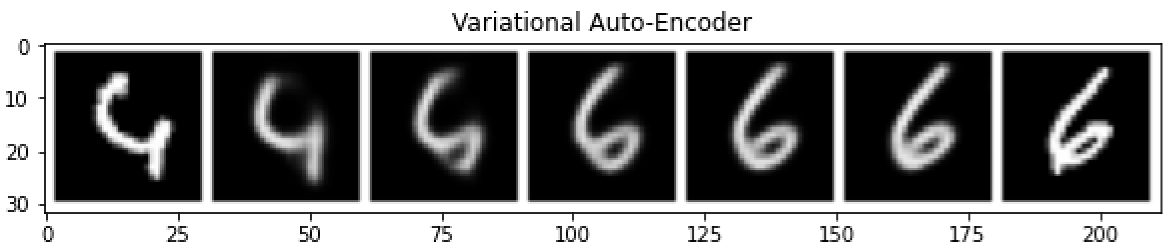
3) Characterize what properties you would expect for reconstructions (1pt) and samples (2pt) of a well-tuned VAE! [3pt]:-

A well-tuned VAE is expected to produce reconstructed samples that will look exactly like the original images with very low blurriness level. In case of the samples, a well-tuned VAE would tackle the problem of the latent space irregularity by making the encoder return a latent space that follows the prior distribution, enabling it to generate images resembling the original ones. The generated samples are not blurry because the beta value is properly tuned in order to maintain the reconstruction quality along with the latent distribution constraint. We can see that the model performs better for the beta value= 0.001 and after some iterations, the KL loss saturates, unlike with Beta=0, where it is always increasing.

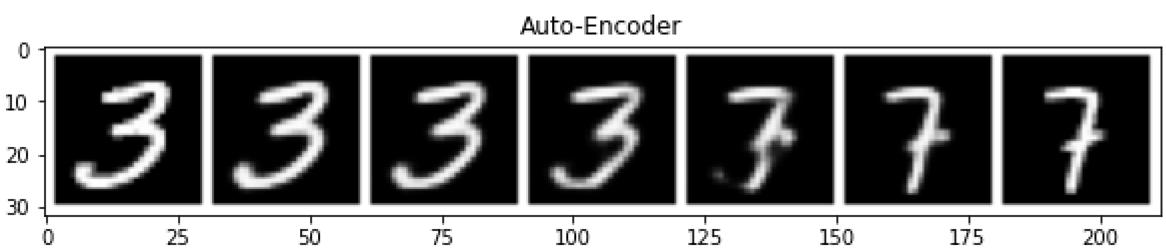
Three representative interpolation comparisons that show AE and VAE embedding interpolation between the same images -

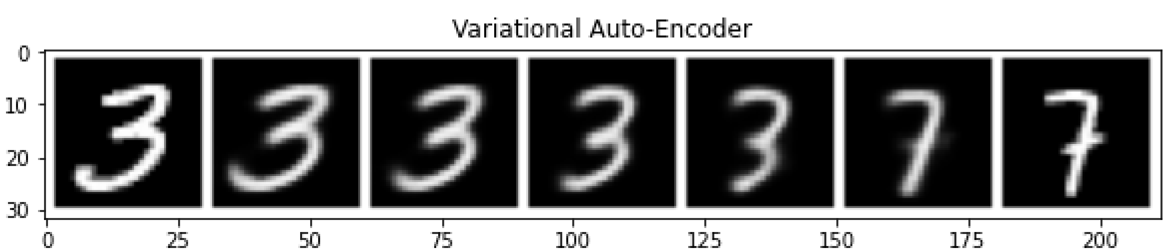
a) 4 and 6:



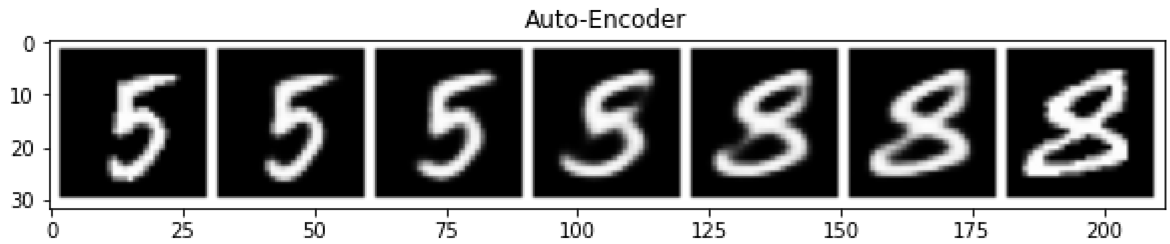


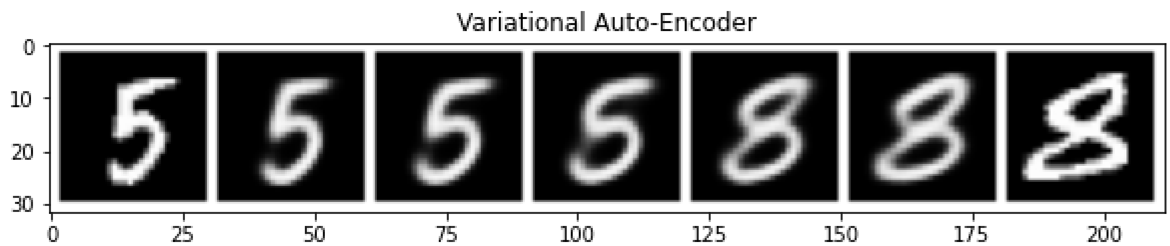
b) 3 and 7:





c) 5 and 8:





Answer to interpolation inline question -

1) We can clearly observe that the AE and VAE embedding space interpolations are different. The AE embedding moves from the start image to the end image, but the edges of the digits are not that prominent. The edges around the digit are not continuous and broken in some steps. But, the VAE embeddings are better with continuous edges and the transition is very smooth from one image to the other.

2) We can see that VAE is performing better than AE. These differences can affect the usefulness of the learned representation for downstream learning, because the representations that are obtained are stable and thus when these representations are used in downstream tasks such as feeding these representations into another model, the output obtained would be more reliable.