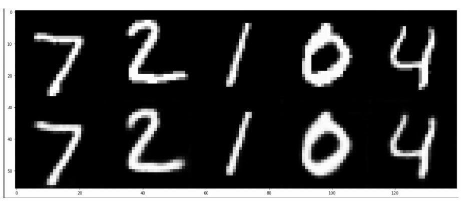
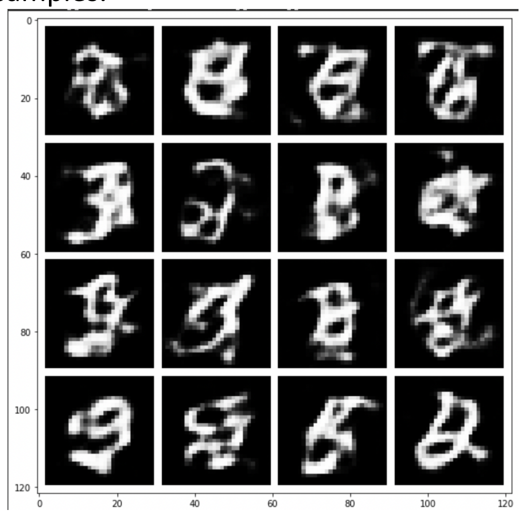
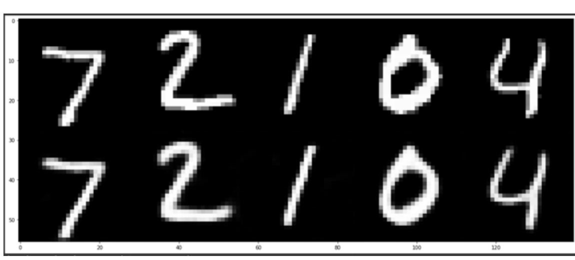
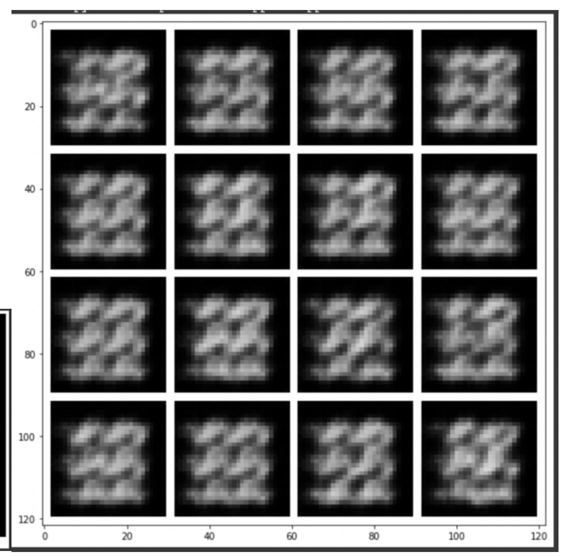
1. We notice that Normal Distribution embeddings produce output image which are convoluted/complex when compared to the actual digits which it has been trained on. This is mainly due to entanglement of latent spaces leading to highly specialized learning of certain shapes, intersections, other patterns. Our output images correspond to the combination of the activations. This is why we have certain decoded images which have complex blurry shapes and are highly complex.
   Below are the reconstructions and samples.

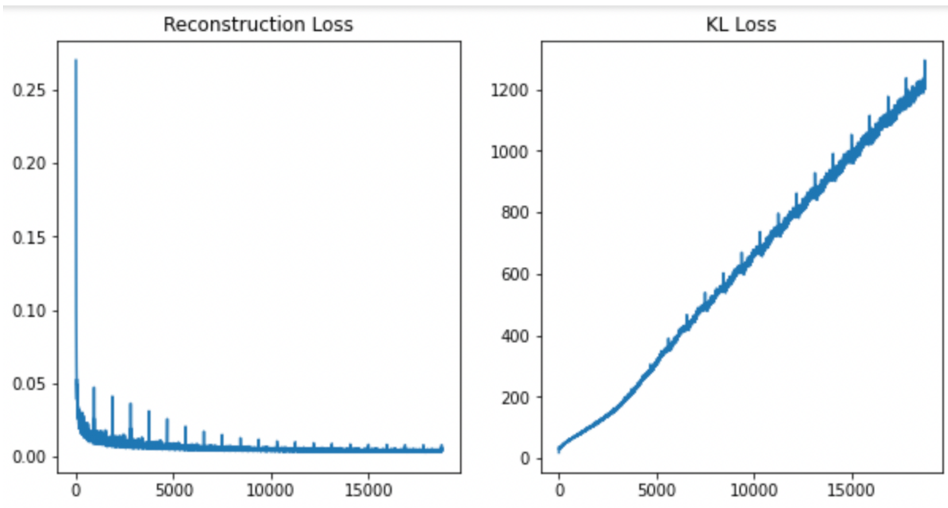




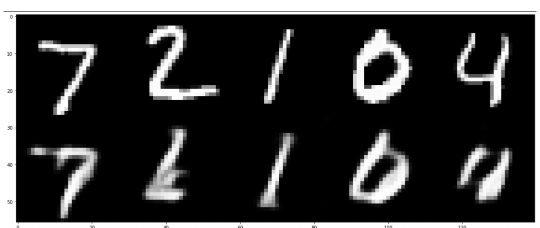
2. 2.1 When β=0,

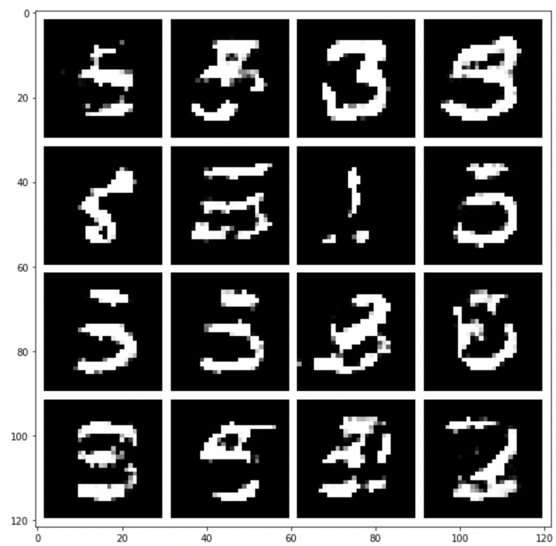


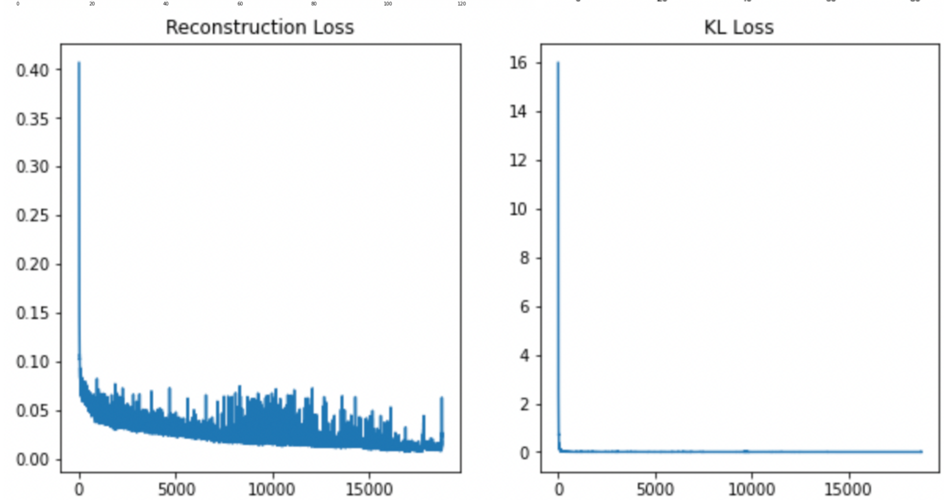




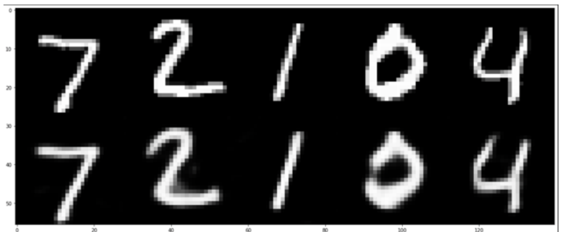
2.2 When β=10

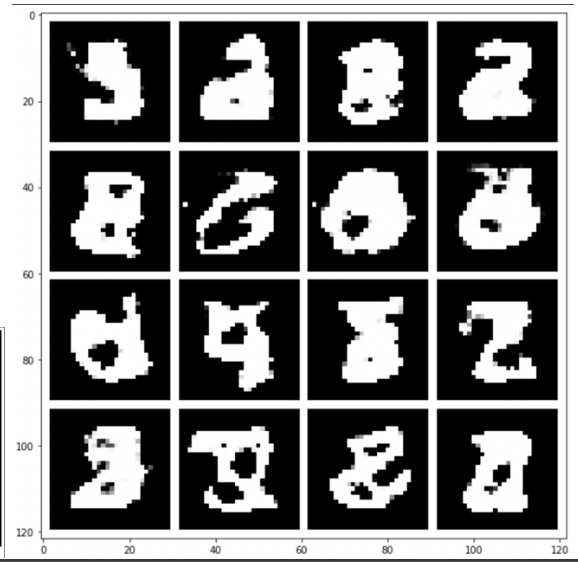


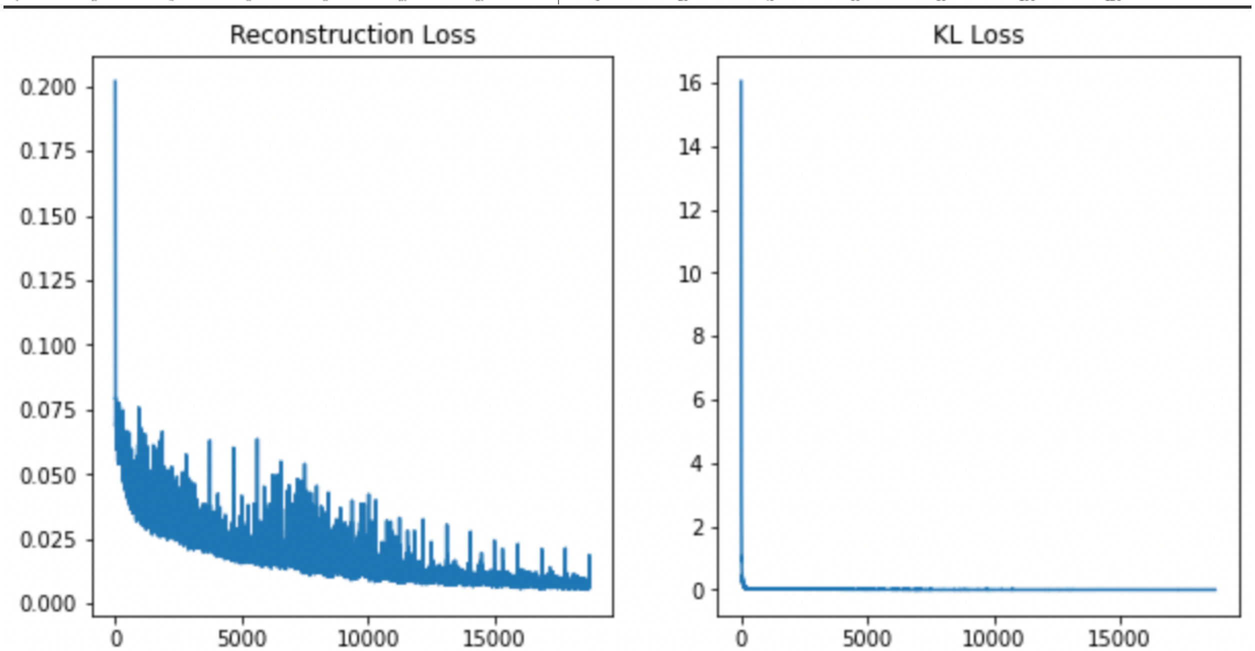




2.3 When β=2.4,







3.

3.1 When β=0, we get the following reconstruction and deconstructed images.

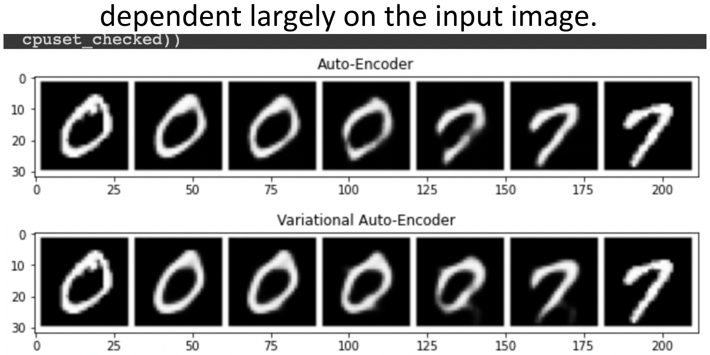
We see near-perfect reconstructions since our losses are directly equivalent to reconstruction loss. The decoded output for the samples are highly convoluted since the regularization term is set to 0. Due to this the VAE learns almost as a traditional autoencoder and only tries to minimize the reconstruction loss. Due to this, the model does not try to recognize distributions/patterns and output high variance activations which in turn lead to the decoded images.

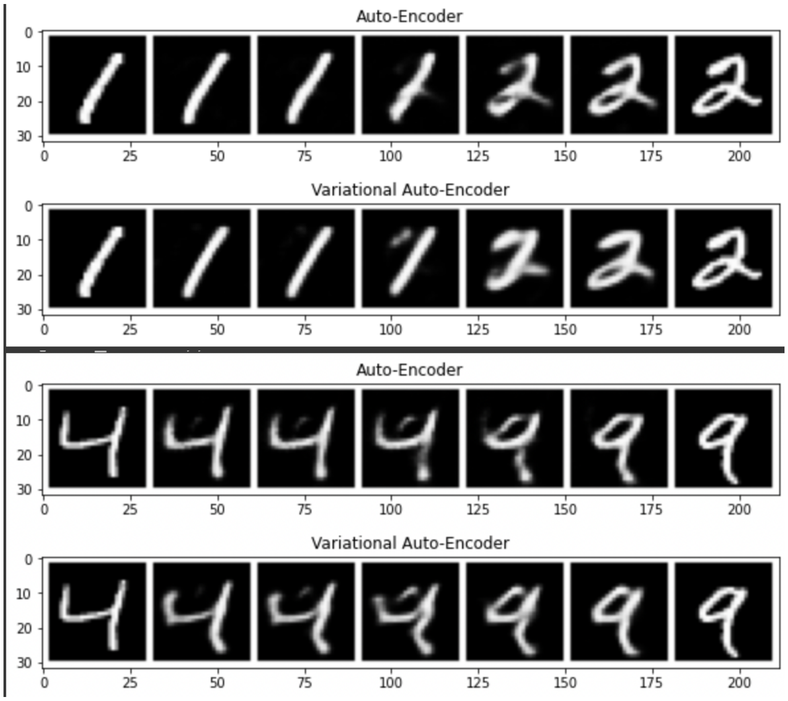
3.2 2.2 When β=10, we get the following reconstruction and deconstructed images.

We see a drop in reconstruction accuracy due to the huge regularization term applied. This leads to some loss of information when deconstructing images. For example, ‘0’ cannot be identified completely due to some ambiguities between ‘0’, ‘6’,’8’ bottom halves. Easier patterns are recognized such as strokes of ‘7’ and ‘1’. Out decoded output for normal distribution outputs less convoluted images since the regularization term allows for disentangled latent spaces and learning of meaningful distributions. Also, because of disentanglement, out decoded outputs are very close to meaningful digits although they are not very clear

3.3 When β=2.4, we get the following reconstruction and deconstructed images. We see accurate reconstructions and decoded images which are close to actual digits and are not as convoluted as the ones with beta=0. Since we have an adequate regularization terms, we are able to decode meaningful digit-like outputs. Samples should be meaningful in terms of patterns and complexities. Also, reconstruction should not be pixel-to-pixel but rather distribution based. The model should behave like a generative model for the samples and generate outputs close to actual number-structure.

4. AE and VAE interpolations differ in the terms of actual reconstruction performed and embedded space representations. VAE learns embedded space with disentangled latent space and therefore can reconstruct images based on the actual output label. Any minor noise in the input image would be lost as the VAE learns concepts(distribution) and reconstructions are more generic. AE on the other hand tries to do a pixel compression and reconstruction, hence they try to preserve each pixel. Output images are hence dependent largely on the input image.





5. VAE works better as a generative model. Downstream models can utilize VAEs as a more generic interpolation and can use outputs as a denoised version of the image. The embeddings learned are also more meaningful and can be utilized. Learned representations can be used in downstream tasks such as labelling, classification, etc. Both can be used for dimensionality reduction and using it for downstream tasks. The latent representation for VAE will likely be more interpretable in terms of human differentiators. We can have downstream tasks based on these embeddings.