

# Assignment 6

## Restricted Boltzmann Machines

### CS7015 : Deep Learning

Siddharth Bhatia (EE16B120)

Anant Shah (EE16B105)

26 April 2019

## 1 Hidden Variables

We can observe from the plots that increasing the number of hidden variables reduces the discrimination power of the clusters in the t-SNE plots.

Lesser number of hidden variables correspond to more compression, leading to the hidden variables capturing only the most important features of the input data. In the case of 10 hidden variables, we can expect to see around 10 clusters, roughly corresponding to the 10 classes.

We have also experimented with the extreme case of 784 hidden variables, which indeed reveals poor separation of clusters. The figures 1 to 6 show the variations in clustering as we alter the number of hidden variables.

## 2 Steps in Gibbs Chain

For the experiment with varying  $k$ , we could ideally expect the performance (like better clustering in tSNE plots or better quality of samples) to increase with increasing  $k$ , since it could potentially get closer to the true distribution before sampling. In other words, it could produce more accurate samples leading to better training.

This is not clearly visible from the t-SNE plots though. The figures 8-12 show the t-SNE plots for the variation in the number of steps in the gibbs chain. The number of hidden variables in this case was  $H = 100$ .

We also compared the re-construction loss for different number of steps in the gibbs chain. The loss plot is shown in Figure 7. We observe that the reconstruction loss decreases over the iterations. The loss has been calculated at every  $1000^t h$  iteration of stochastic gradient descent. We observe that the loss decreases over SGD iterations which means that the samples being obtained from the Gibbs chain are close to the initial image we started off with. The reconstruction loss is almost same for different steps in the gibbs chain. However finally,  $k = 1$  provides us with the least reconstruction loss.

## 3 Samples from Contrastive Divergence

The samples generated by Gibbs chain after every 100 steps of SGD are shown in Figure 12. We can see that as the model trains, the samples obtained from the gibbs chain become closer to that of the initial sample (training image). Figure 14 shows the comparison of the initial sample and the sample obtained from the gibbs chain for  $k = 1$  after running SGD for around 100,000 iterations.

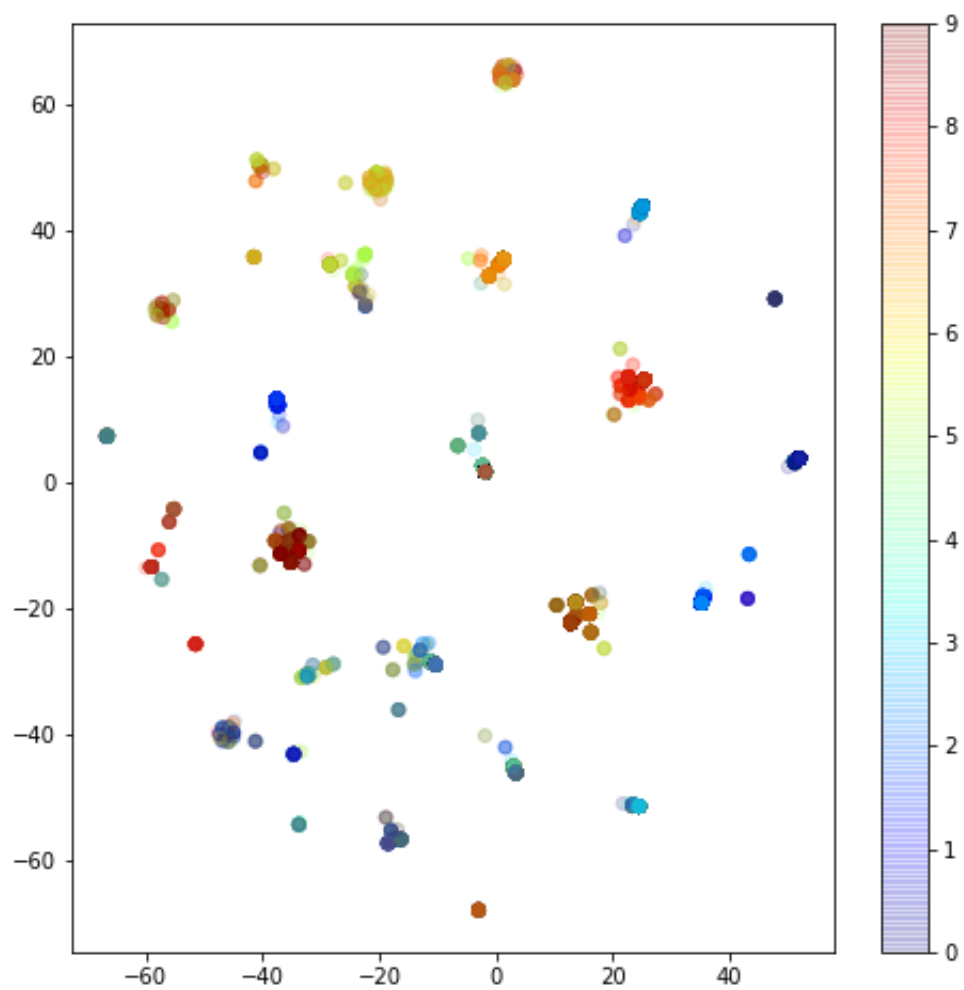


Figure 1: 10 Hidden Variables

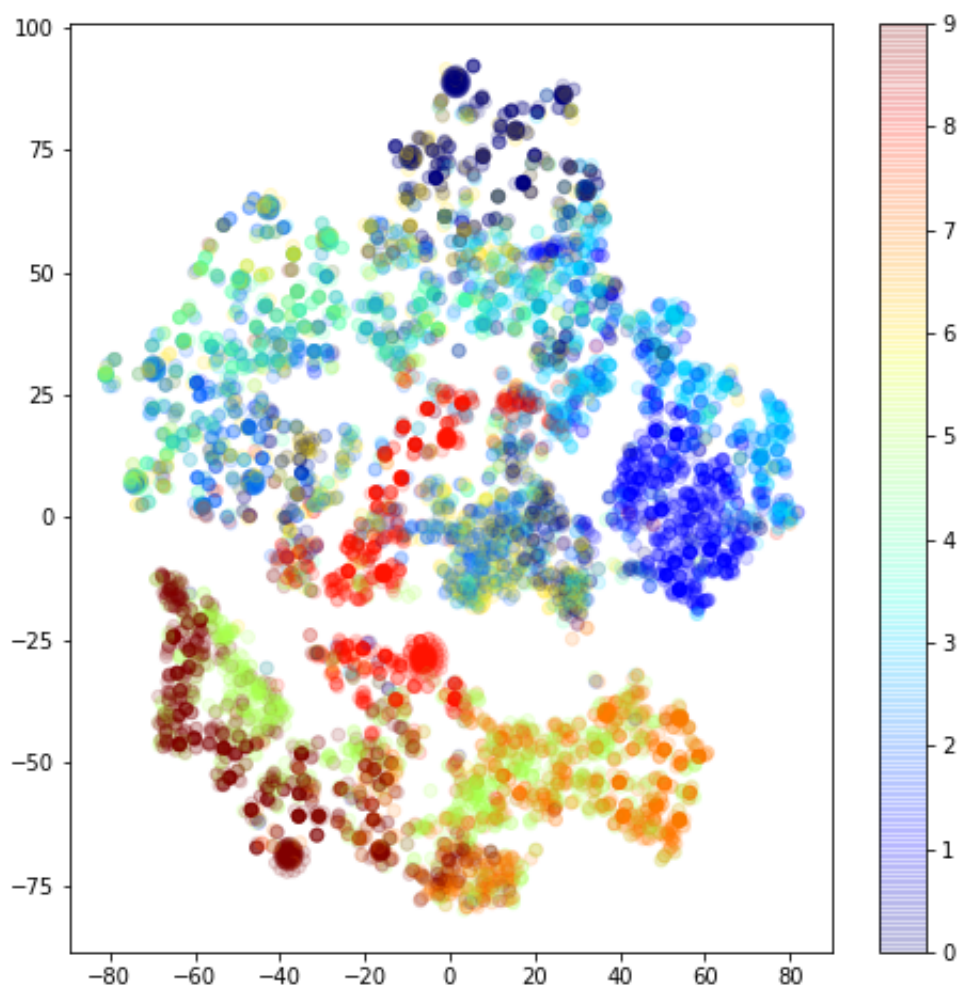


Figure 2: 50 Hidden Variables

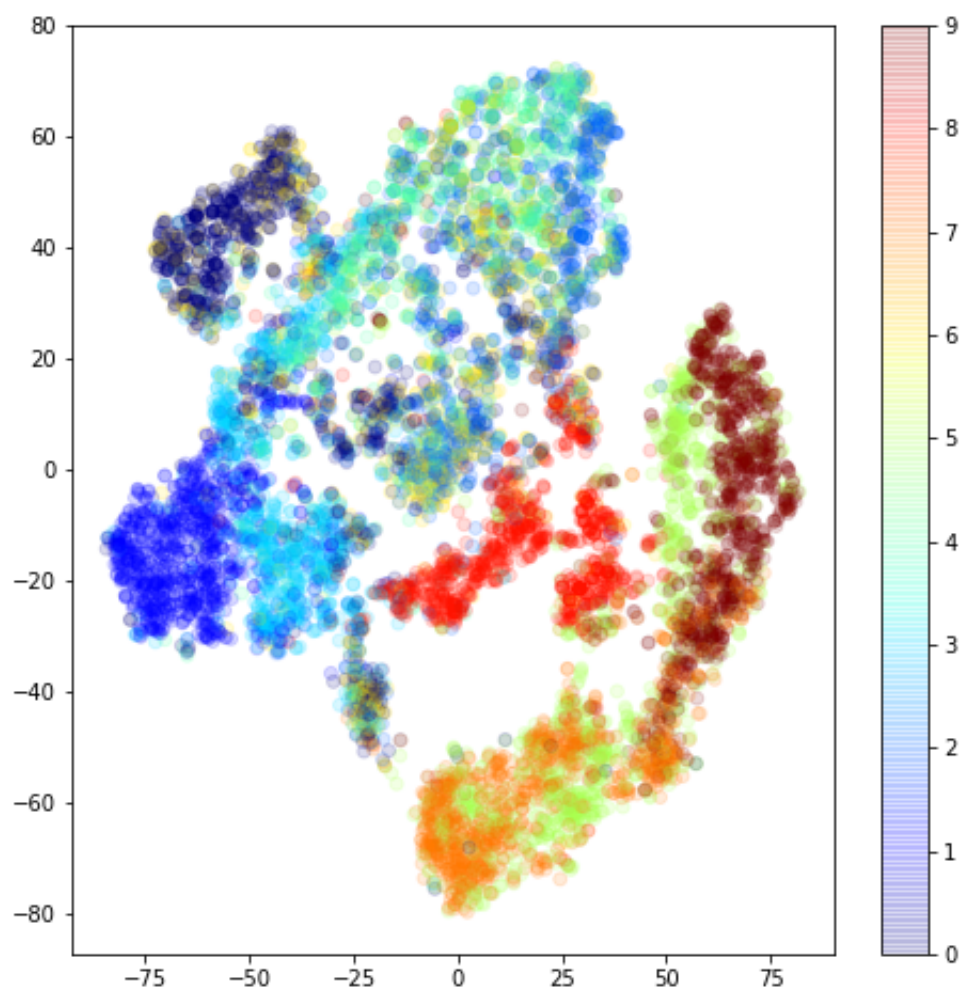


Figure 3: 100 Hidden Variables

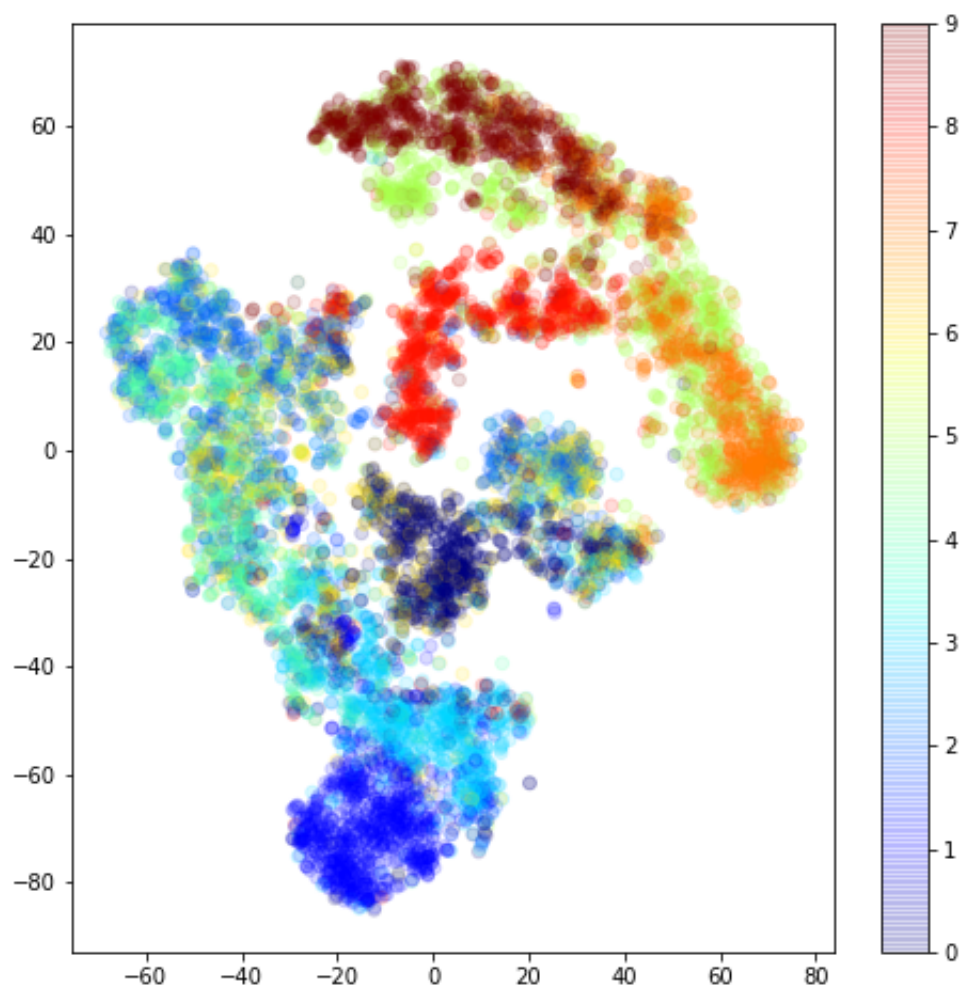


Figure 4: 200 Hidden Variables

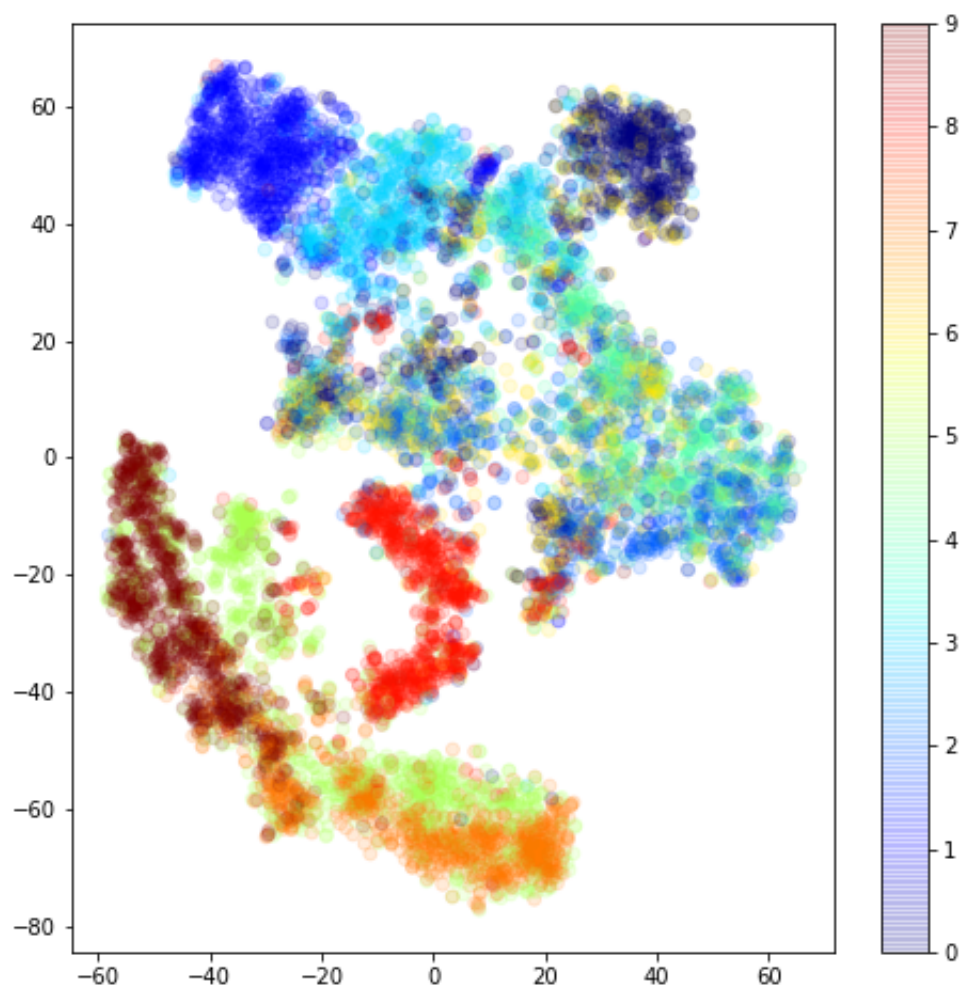


Figure 5: 300 Hidden Variables

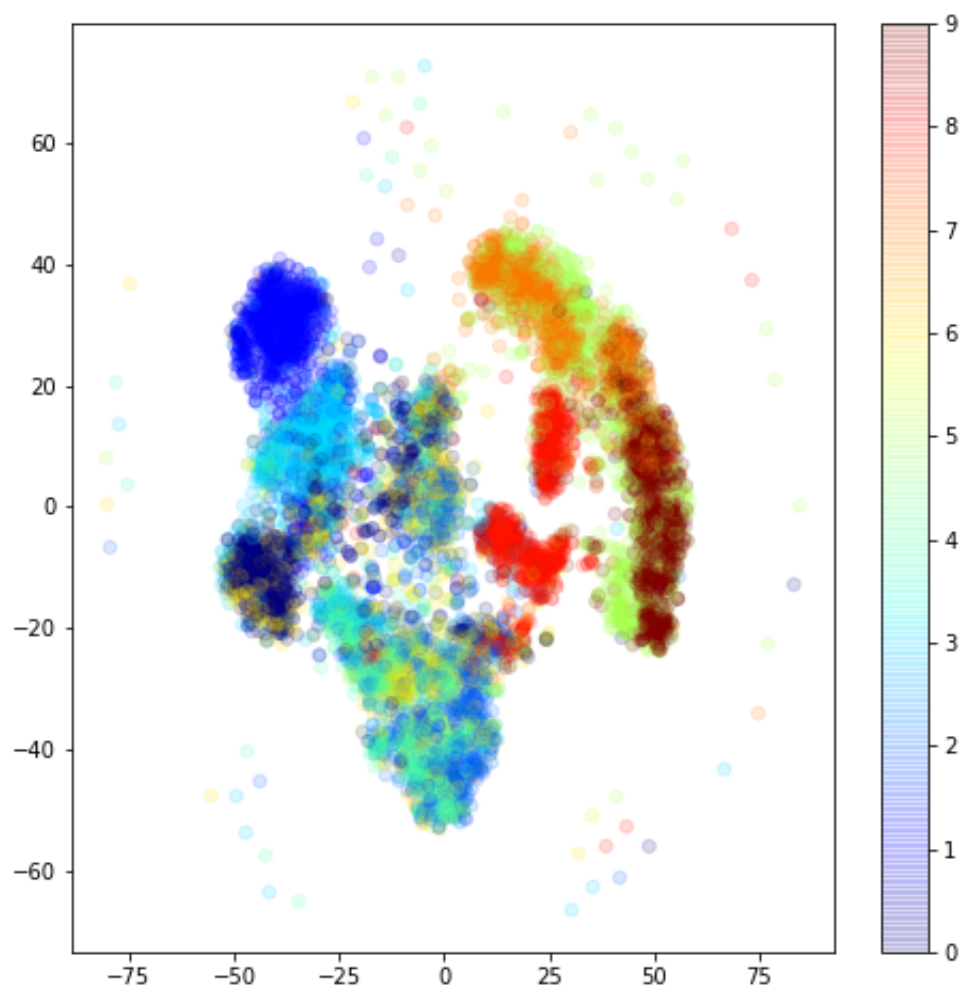


Figure 6: 784 Hidden Variables

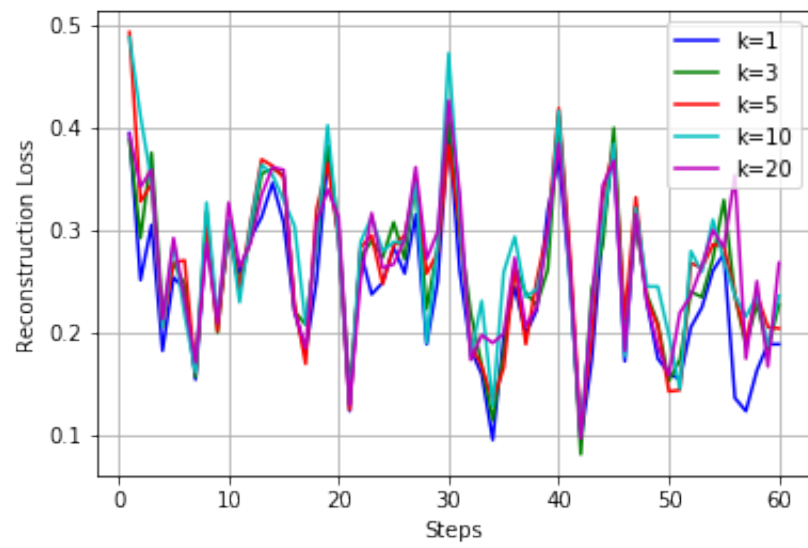


Figure 7: Reconstruction Loss



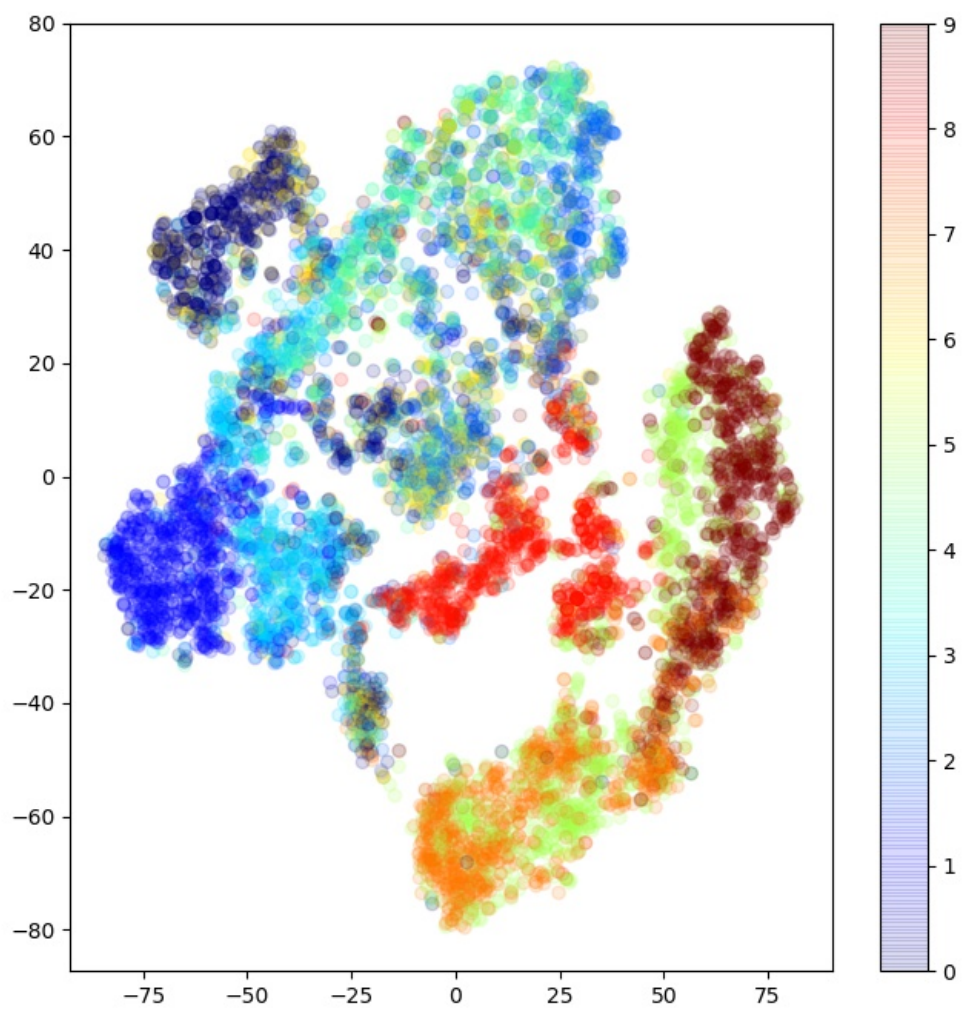


Figure 8:  $k=1$

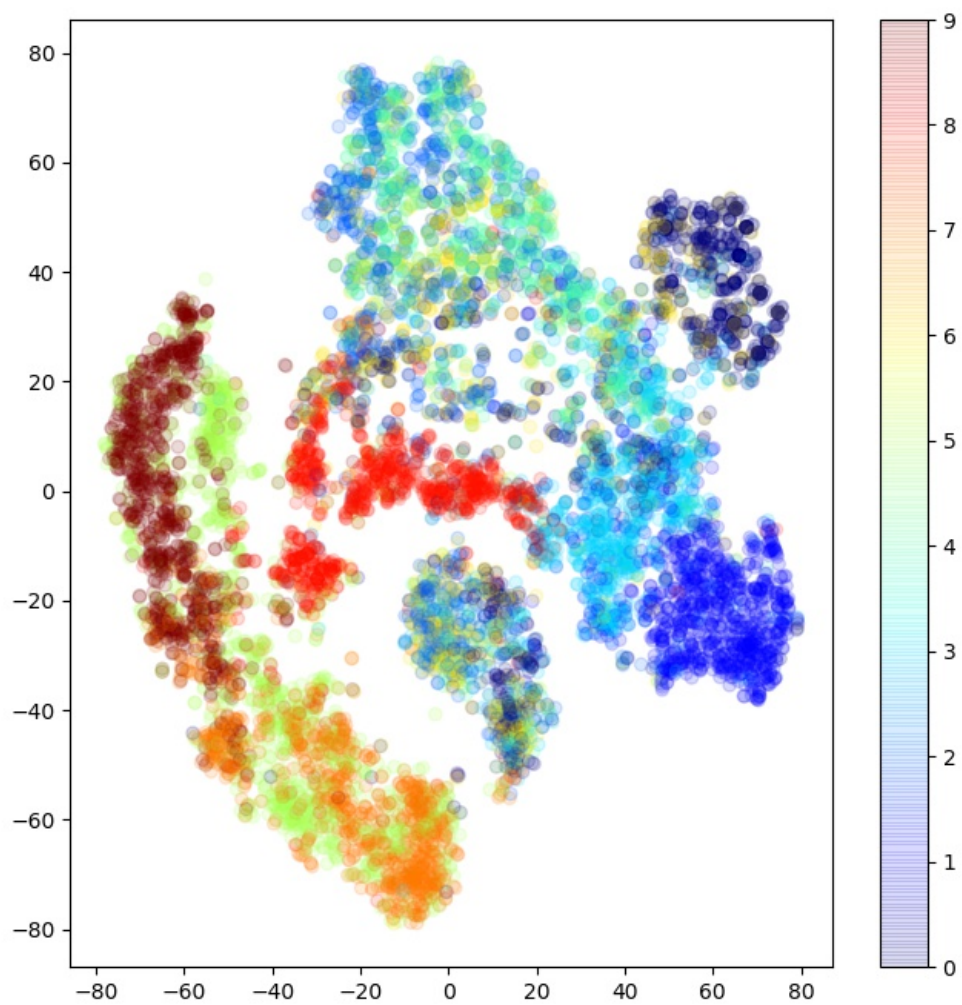


Figure 9:  $k=3$

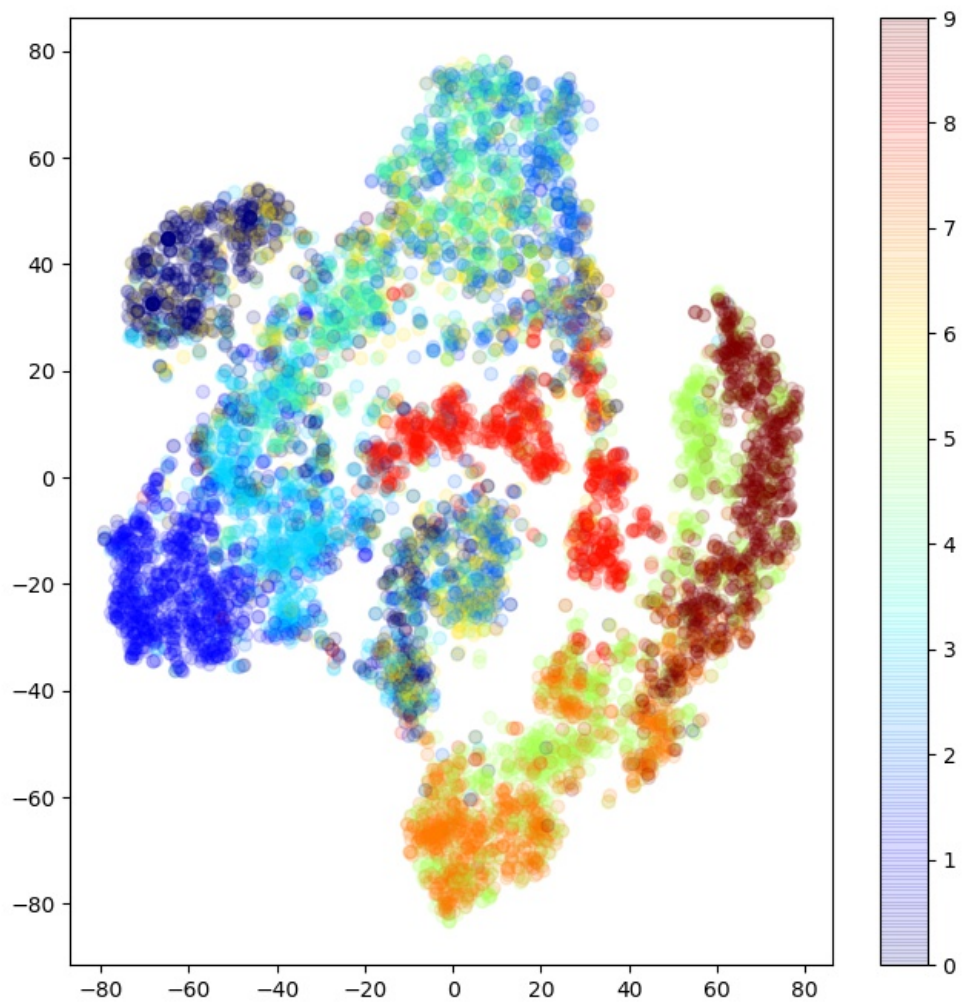


Figure 10:  $k=5$

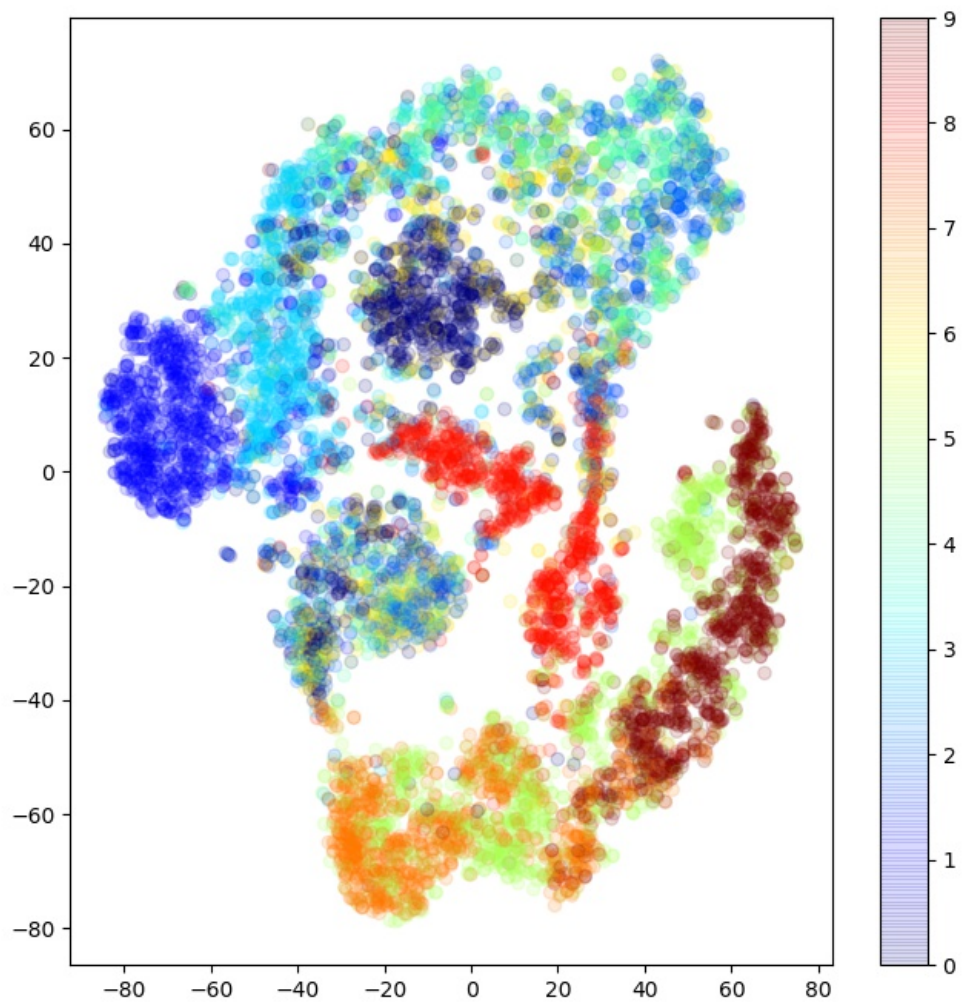


Figure 11:  $k=10$

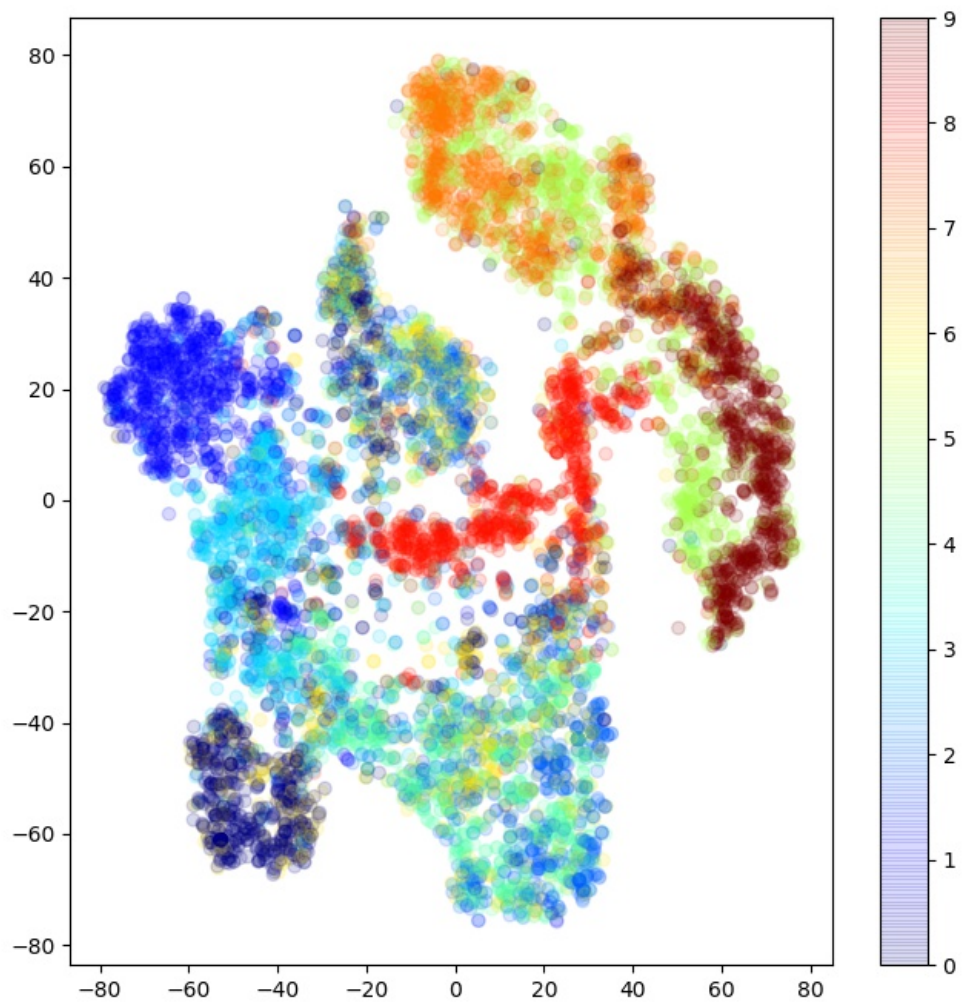


Figure 12:  $k=20$

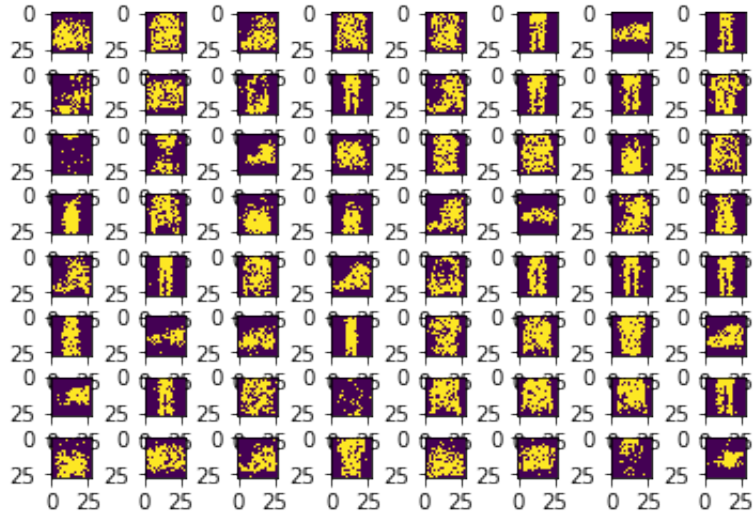


Figure 13: Samples from Contrastive Divergence

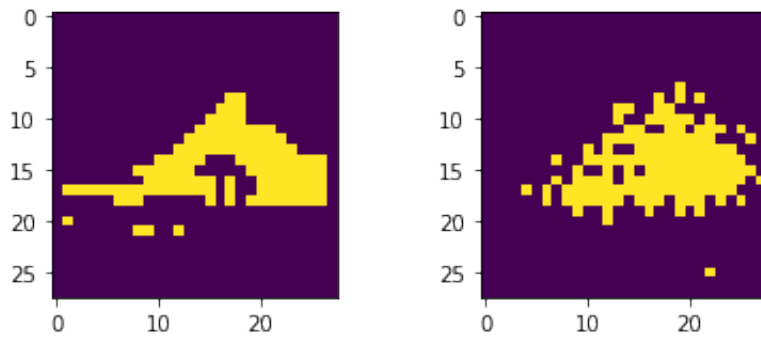


Figure 14: Comparison of initial sample and generated sample