Report for Programming Assignment 4

April 2019

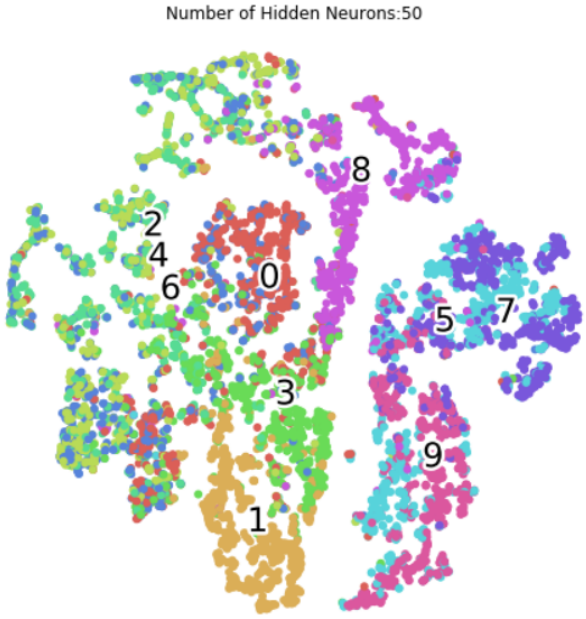
Team Mate 1 : Anshul Kumar
Roll No : ME15B082
Team Mate 2 : Ashutosh Raj
Roll No : ME15B086

1 2-D t-SNE plots for learned representations

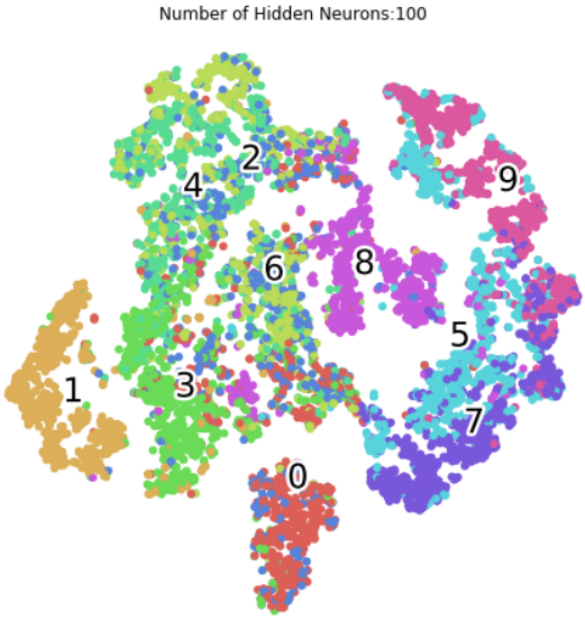
Test data hidden representations have been plotted for various values of n (number of dimensions of hidden representation).

Configuration followed:
Learning Rate = 0.001
Batch_size = 1 (SGD)
Epochs = 1
k (Gibbs chain steps) = 1

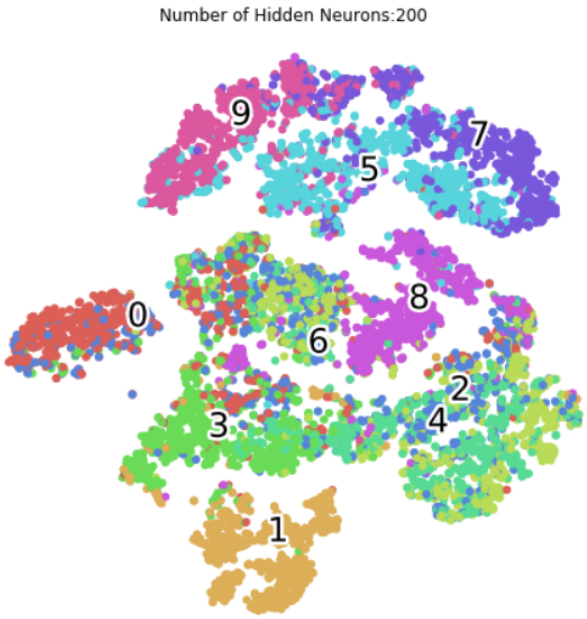
Number of hidden representations (n) = 50



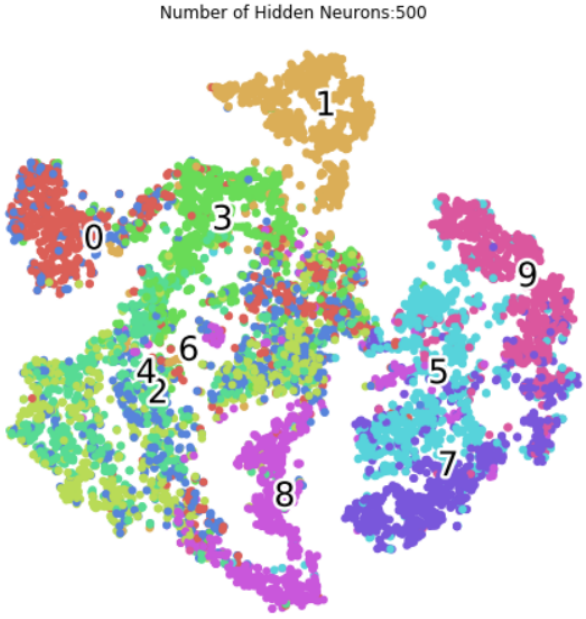
Number of hidden representations (n) = 100



Number of hidden representations (n) = 200



Number of hidden representations (n) = 500



Closer clusters might show similar learning characteristics, thus similar classes are expected to be closer

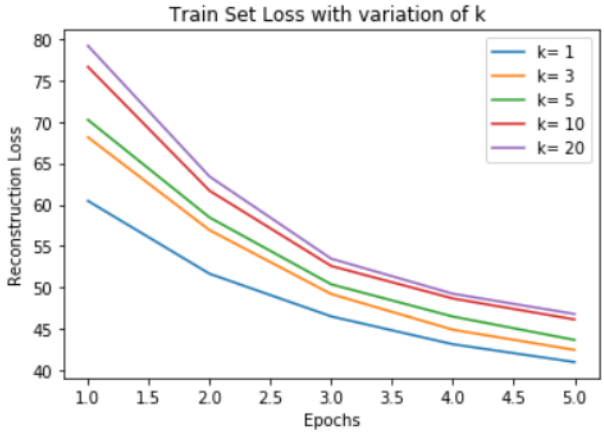
With increase in the value of n, training loss decreases and thus, we have fairly separated clusters with large n. For n =50, clusters are poorly separated while for n larger than 100, clusters show good separation.

Although, with increase in n, training and t-SNE plot formation time also increases.

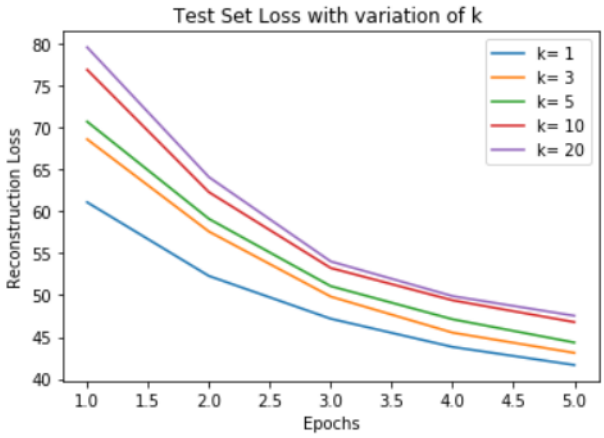
2 Varying Gibbs Chain: k

Configuration followed:
Learning Rate = 0.001
Batch_size = 1 (SGD)
Epochs = 5
Number of hidden representations (n) = 100

Train Set Loss

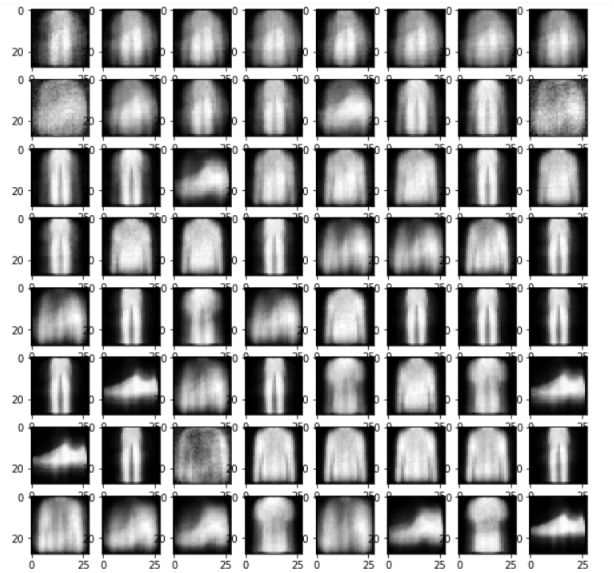


Test Set Loss



We observe that as we increase the value of k (i.e. increase the Gibbs Chain), the loss curve is seen to be below that of the curve corresponding to lower k. CD (with k = 1) may be a bad approximator of the log-likelihood but it ensures that hidden features retain most of the information in the data vector, while CD with higher k might be a better approximation to log-likelihood but not be good at retaining the information in the data vector. Hence, while training the RBM, we see that the error is more when K is increased.

3 Samples generated after Gibbs Chain after every 100 iterations



Initially, the samples obtained carried large noise, but as the training progresses samples become smoother and smoother.