CS7015-Deep Learning

Programming Assignment#5:RBM Report

Satish G(CS14B042)
Ajmeera Balaji Naik(CS14B034)

April 29, 2018

---

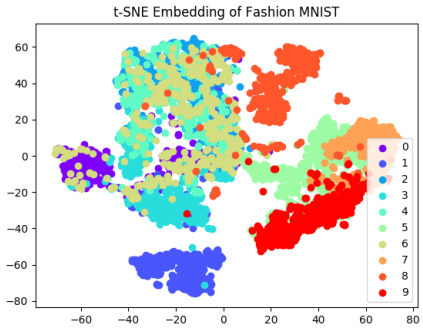
1 Experimentation Details:

- The below parameters settings are used for experimentation
  - *Learning Rate* - 0.01
  - *Hidden units* - 400,500,600,700
  - k - 1, 5, 10
- The weights are initialized using a **Gaussian** with $\mu = 0$ $\sigma = 0.01$ and the biases are initialized to *zero*
- Loss: *Reconstruction error*
  - $\frac{\sum_{d=1}^{N}(V_d - \tilde{V_d})^2}{N*784}$ where V_d is the original data point, $\tilde{V_d}$ is the reconstructed(sampled) from the model.
- The t-SNE plots are plotted for validation data-set

2 PS Questions:

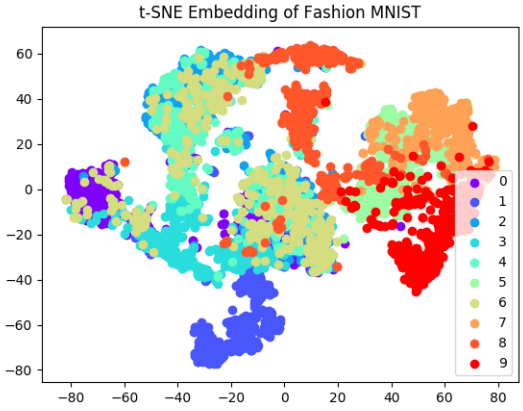
1. Use t-SNE to plot the learned representations in a 2-dimensional space (t- SNE will essentially take the n-dimensional representation and plot it in a 2d space). While plotting use a different color for each of the 10 classes and see if you see interesting clusters. Experiment with different values of n

t-SNE plot for original(dim=784) data(valid) points:

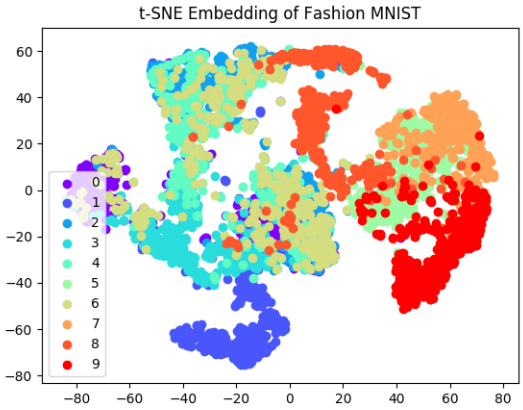


t-SNE plots for different values of n:

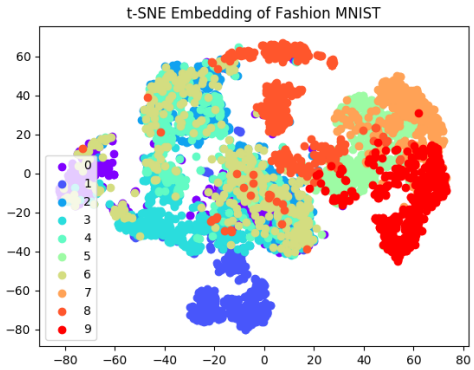
- For $n = 400$ and $k = 1$:



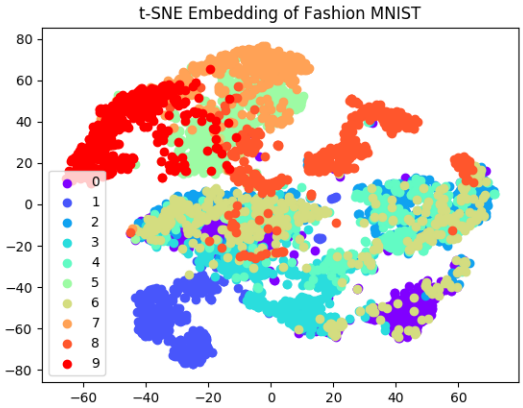
- For $n = 500$ and $k = 1$:



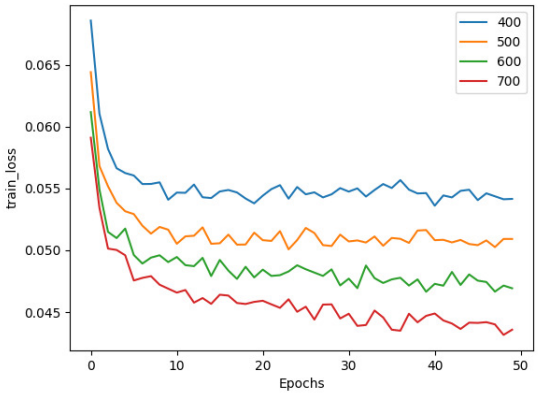
- For $n = 600$ and $k = 1$:

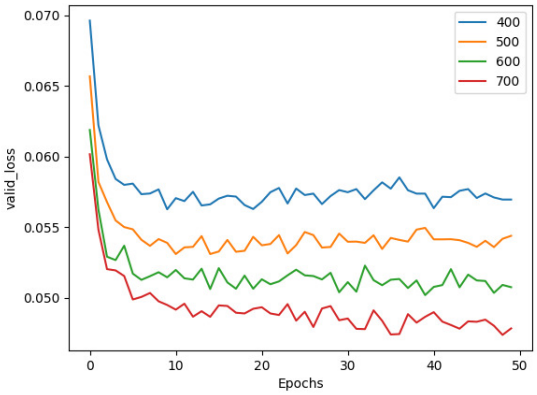


- For $n = 700$ and $k = 1$:



Train & Valid loss for different values n:





Observations:

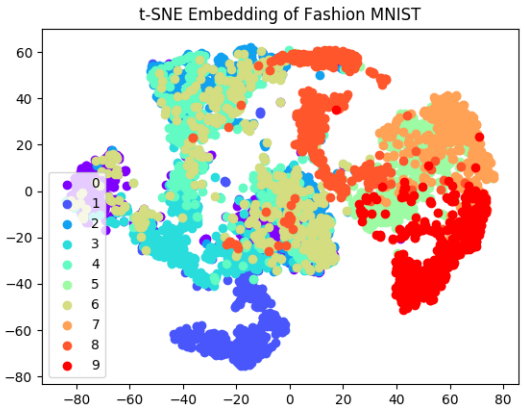
- It is observed that both the training and validation errors were decreasing as number epochs increased.
- As n which is maximum dimension of hidden representation is increased, the training and validation errors were kept decreasing at for any given number of epochs.

Inferences:

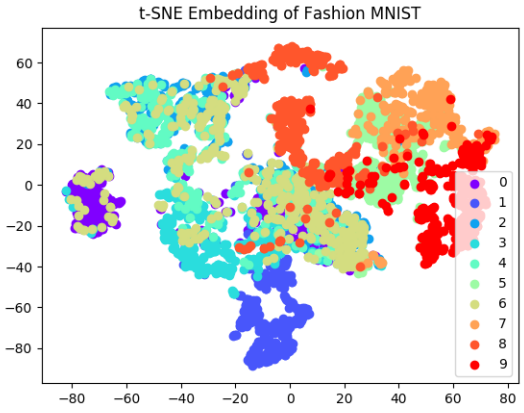
- As number of epochs(for a given k) increased, the system weights are improved such that only relevant and non-redundant information is captured and redundant information was ignored. Hence the reconstructed vectors looked increasingly similar to original vector as number of epochs were run increased and so the reconstruction error(loss) **decreased**.
- As the dimension of hidden representation increased, since training error was decreased, it means system was able to do reconstruction well. We have roughly estimated the entropy of data(with some reasonable assumption). It turns out that nearly 732 i.e these many number of bits required to represent a input vector on average. It also means that to capture the hidden representation perfectly, the dimension of hidden representation should be equal to 732. As we increased the dimension of hidden representation, increasingly good hidden representation was captured.Hence the system was able to do reconstruction well. Hence the error was decreasing.

2. In every step of stochastic gradient descent (SGD) you will be running the Gibbs Chain for k steps. Study the effect of using different values of k.

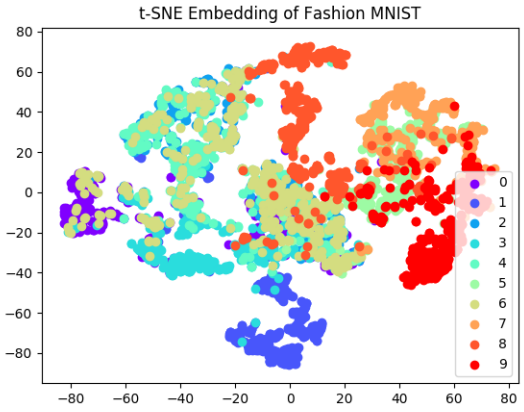
- **t-SNE plot for $n = 500$ and $k = 1$:**



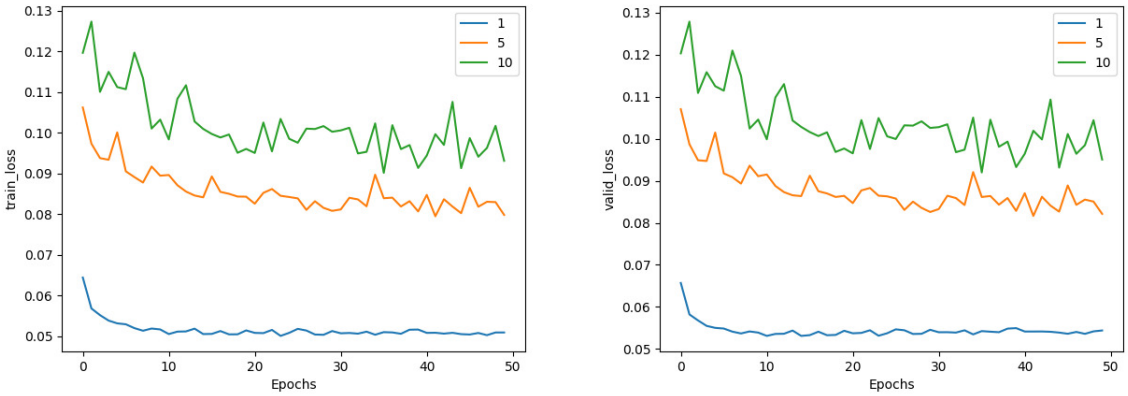
- **t-SNE plot for $n = 500$ and $k = 5$:**



- **t-SNE plot for $n = 500$ and $k = 10$:**



Train & Valid loss for different values k:

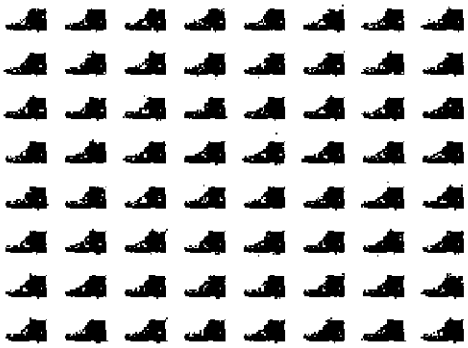


Observations:

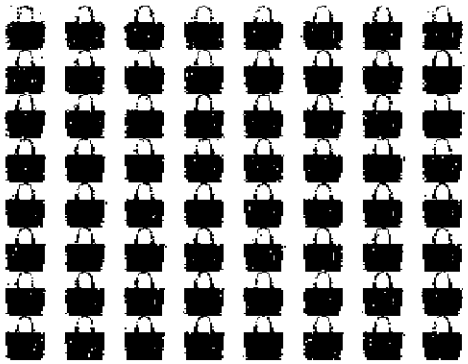
- For given k value, errors were decreased as number of epochs increased.
- For a given number of epochs, errors were high for higher values of k.

3. Suppose CD takes m iterations of SGD to converge. Plot the samples m steps of SGD. Use an 8 8 grid to plot these generated by Gibbs chain after every 64 64 samples.

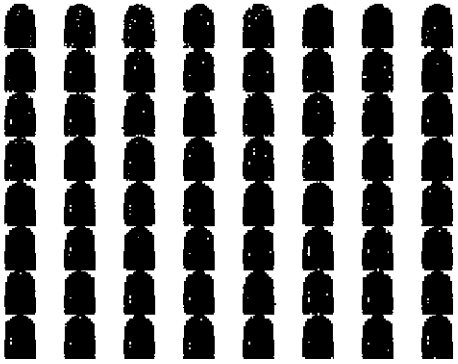
- **Samples of Ankle boot:**



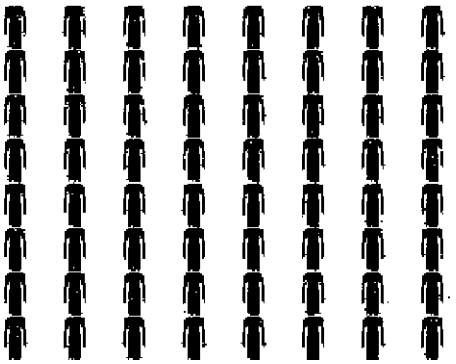
- **Samples of Bag:**



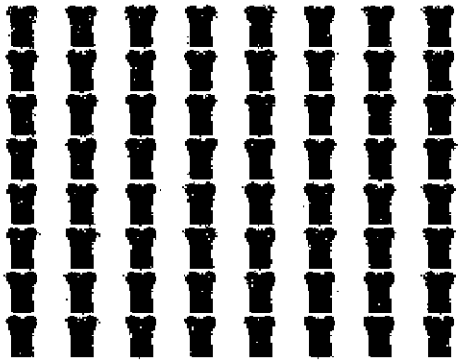
- **Samples of Coat:**



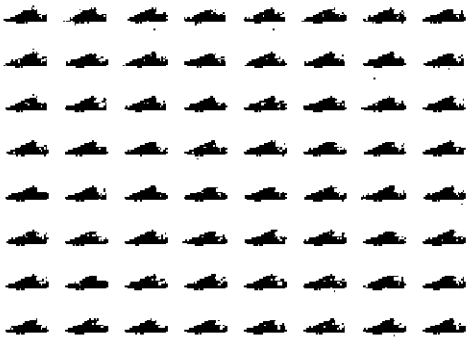
- **Samples of Dress:**



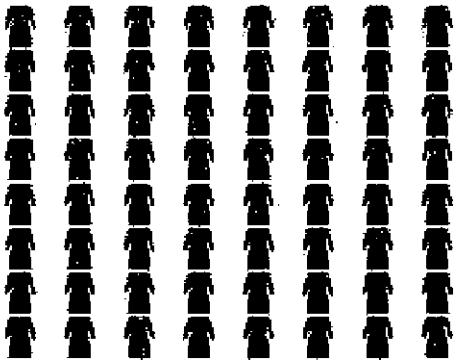
- **Samples of Shirt:**



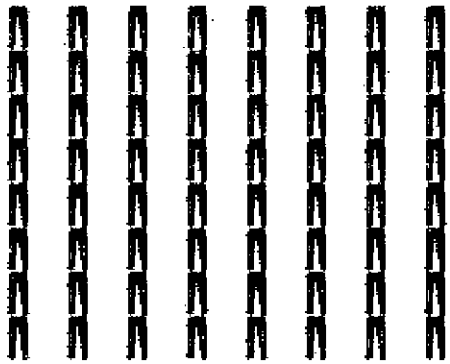
- **Samples of Sneaker:**



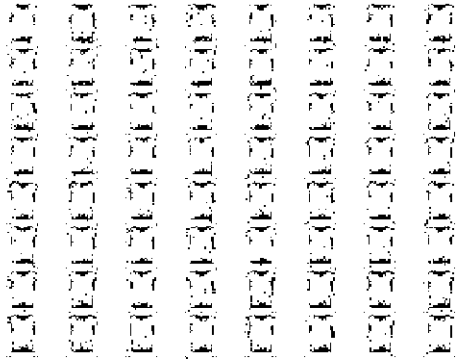
- **Samples of Top:**



- **Samples of Trouser:**



- **Samples of Pullover:**



- **Samples of Sandal:**