Report for Programming Assignment-4

April 25, 2019

Checklist :

☑ checked.

☐ unchecked.

☒ not done.

The task you need to ensure before submission.

- ☑ We have read all the instruction carefully and followed them to our best ability.
- ☑ We have written the name, roll no in report.
- ☑ Run sanity_check.sh.
- ☑ We will be submitting only single submission on behalf of our team.
- ☑ We have not included unnecessary text, pages, logos in the assignment.
- ☑ We have not used any high level APIs(Keras, Estimators for e.g.).
- ☑ We have not copied anything for this assignment.

Team Mate 1 : Ojas Mehta
Roll No : CS18M038
Team Mate 2 : Avani Shukla
Roll No : CS18M052
Team Name : 101010101

Report

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 such the images which are close in the n-dimensional space will be close in the 2d space also). 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.

- Plot of Learned Representation for n = 32

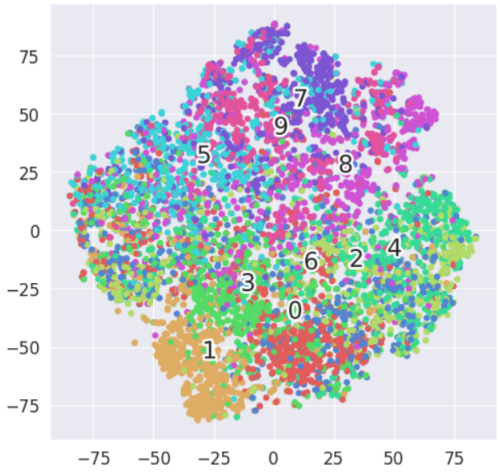

Figure 1: Learned Representation for n = 32

- Plot of Learned Representation for n = 64

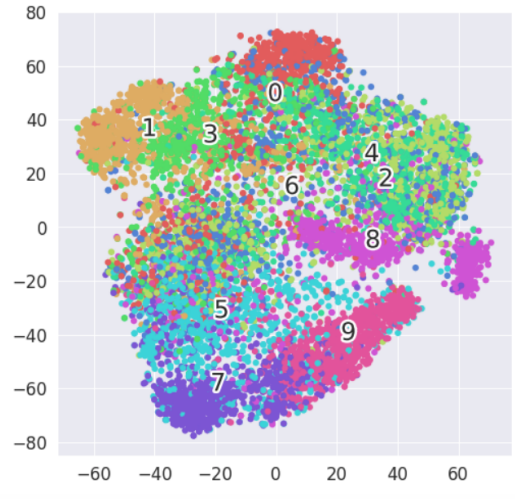

Figure 2: Learned Representation for n = 64

- Plot of Learned Representation for n = 128

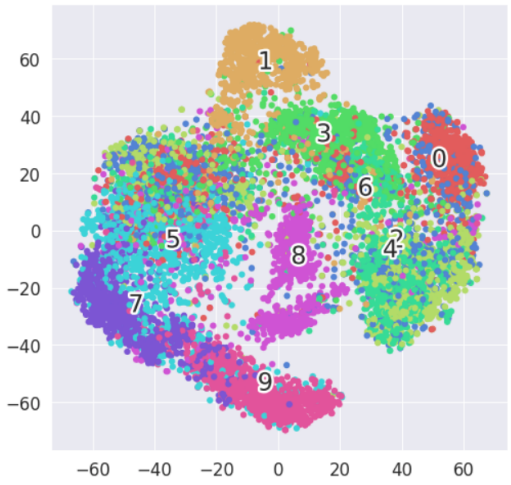

Figure 3: Learned Representation for n = 128

- Plot of Learned Representation for n = 256

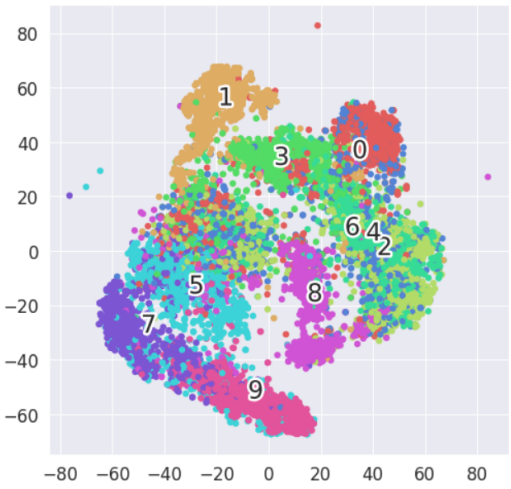

Figure 4: Learned Representation for n = 256

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 : 1, 3, 5, 10, 20.

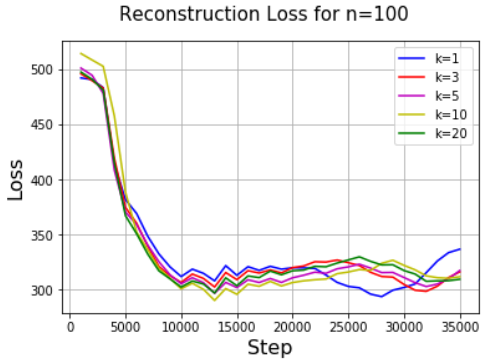

Figure 5: Reconstruction Loss vs step plot for different values of k

3. Assume that CD takes around m iterations of SGD to converge, where m = 6400. Plot the samples generated by Gibbs chain after every $\frac{m}{64}$ steps of SGD. Use an 8×8 grid to plot these 64 samples.

   - CD takes around m = 120000 iterations (2 epochs) of SGD to converge.

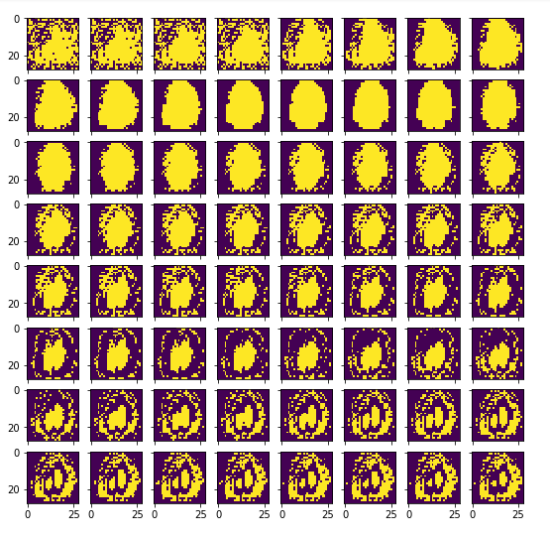

Figure 6: Samples Generated after every 1875 steps of SGD

CONCLUSIONS:

- As the number of hidden units increased, the seperation between classes became more clear. For n = 256, the interclass separation is maximum.

- With regards to Figure 5, for a particular k, loss decreased with increase in number of updates which indicates that training is fine.
  For the first 10000 updates the decrease is loss is much more significant than after it.

- Although, there is not much change in loss across different k but as k increases from 1 to 10, the loss decreased marginally.

- For Figure 6, after each 1875 weight updates the reconstructed image seems to get better and more specific.
  That definitely follows from the model getting trained better and loss decreasing as its already confirmed in Figure 5.