Indian Institute of Technology, Madras

Deep Learning

Course: CS7015

Department of Computer Science and Engineering

DLM

Monika Chauhan
CS17M026

Monika Ahirwar
CS17M025

Restricted Boltzmann machines

Question Answers:

1. Plot of the learned representations in a 2-dimensional space for different values of n.

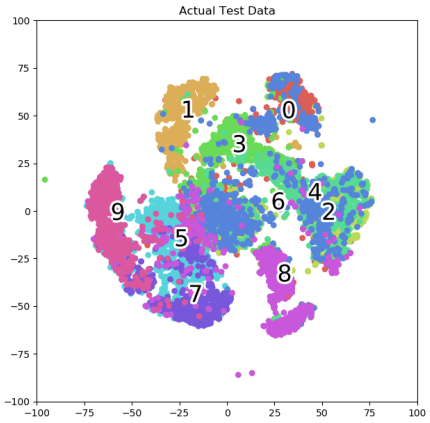
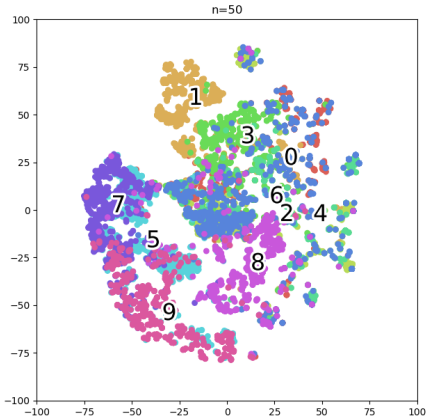
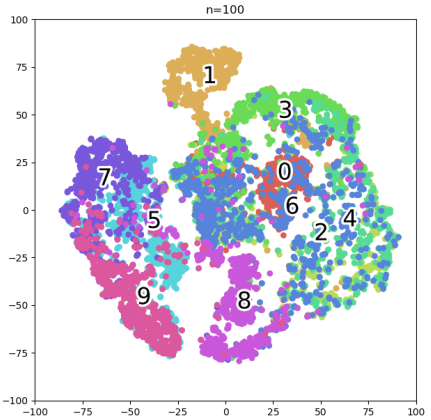


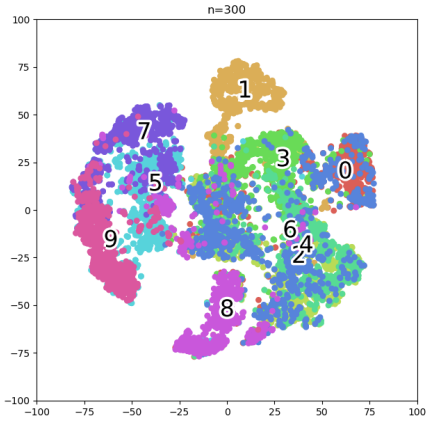
Figure 0.1: Actual Test Data



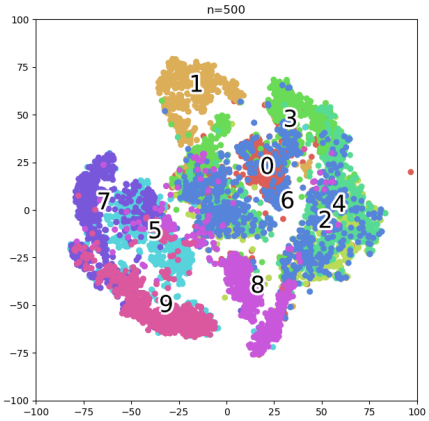
(a) Representation for n=50



(b) Representation for n=100



(a) Representation for n=300



(b) Representation for n=500

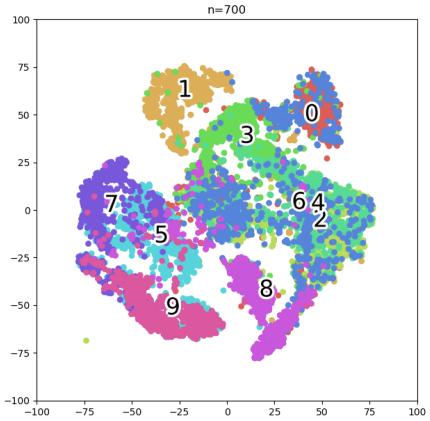


Figure 0.4: Representation for n=700

Other Parameters:

Learning rate: .001
k: 1
Epochs: 10

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.

   Loss decreases faster as k increases

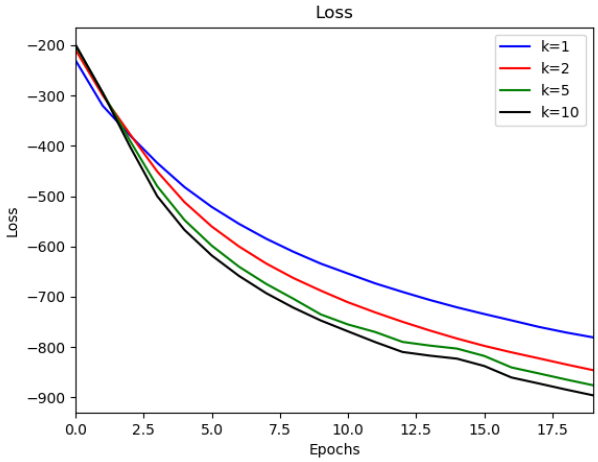
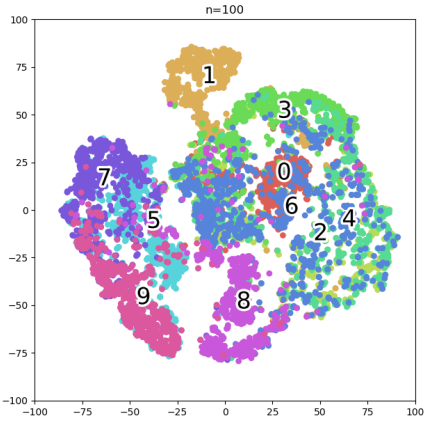
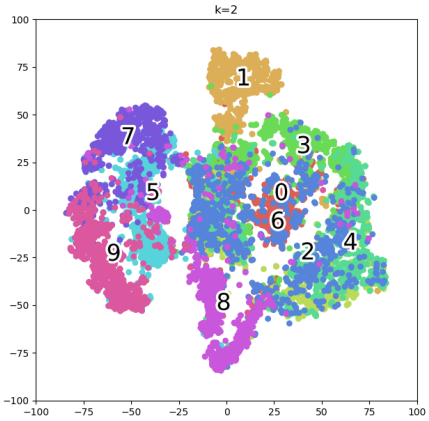


Figure 0.5: Loss for n=700

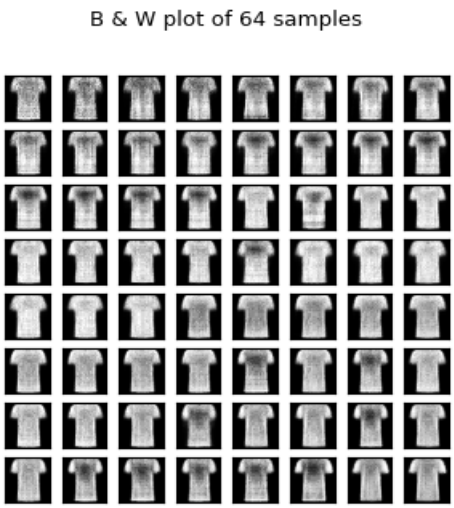


(a) Representation for k=1

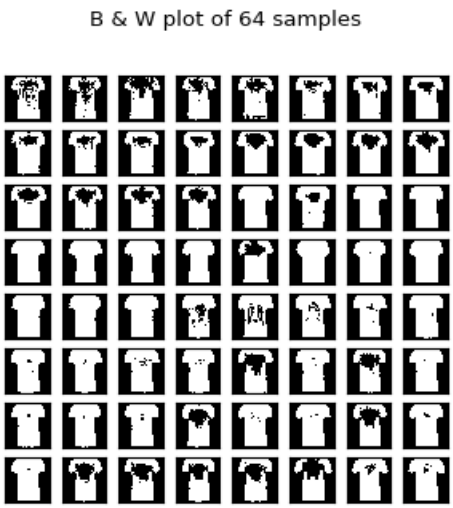


(b) Representation for k=2

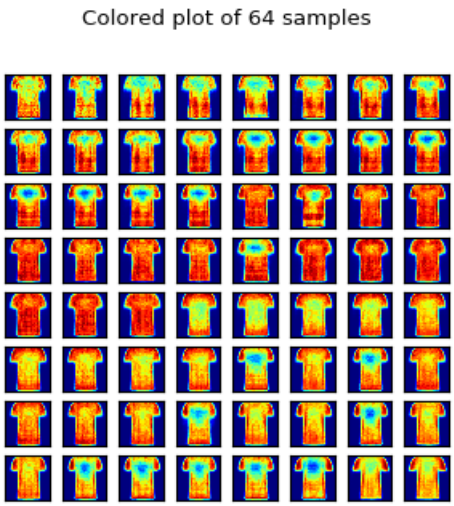
3. Suppose CD takes m iterations of SGD to converge. Plot the samples generated by Gibbs chain after every $\frac{m}{64}$ steps of SGD.



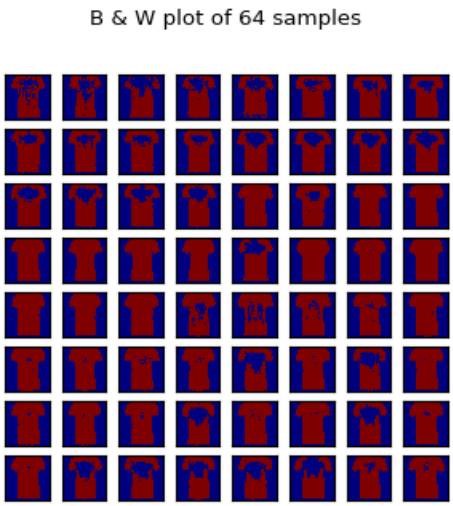
(a) Visible weights



(b) Visible pixels in binary



(a) Visible weights



(b) Visible pixels in binary

Number of updates m is 480,000
Stochastic Gradient Descent has been used
Plots are drawn at every $m/64^{th}$ step for a chosen data point
Plots are also drawn in BnW for better visualization

4. Instead of CD use Gibbs Sampling. How many steps do you need to run the chain for before you start seeing samples from P (V, H)? Does this number change as SGD reaches convergence ?

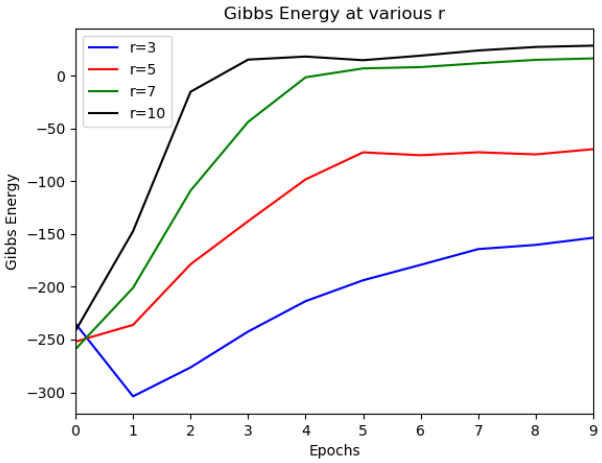


Figure 0.9: Gibb's Energy at K=5 and r=5,7,10

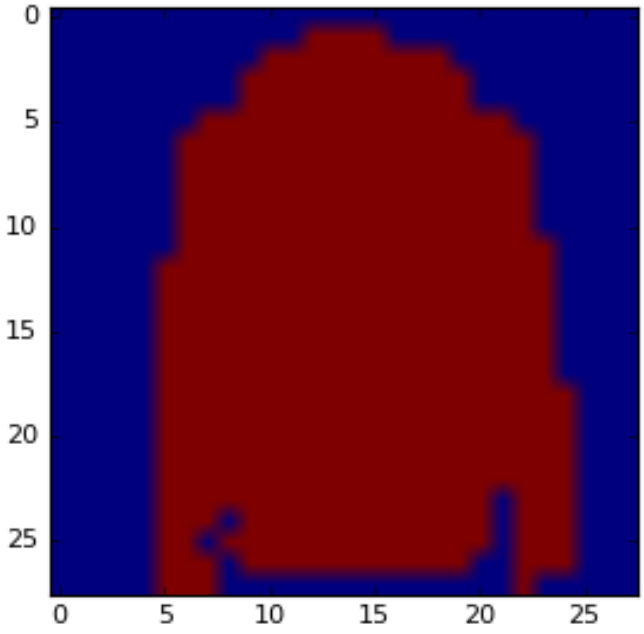
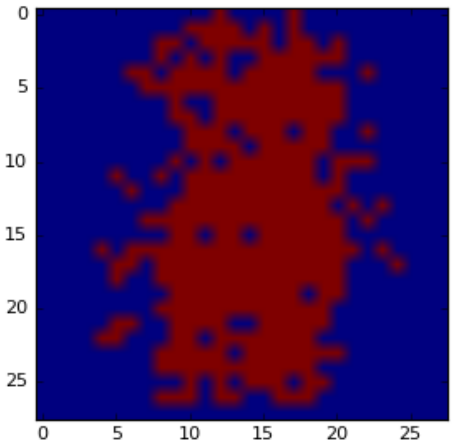
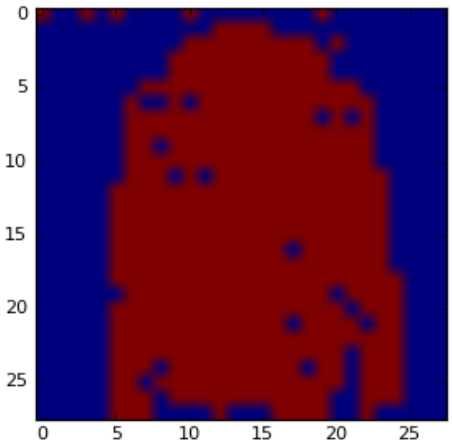


Figure 0.10: Original Image



(a) r=10



(b) 25

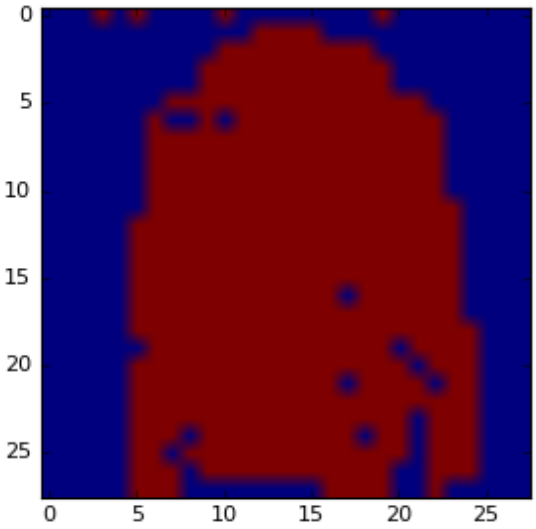


Figure 0.12: Original Image