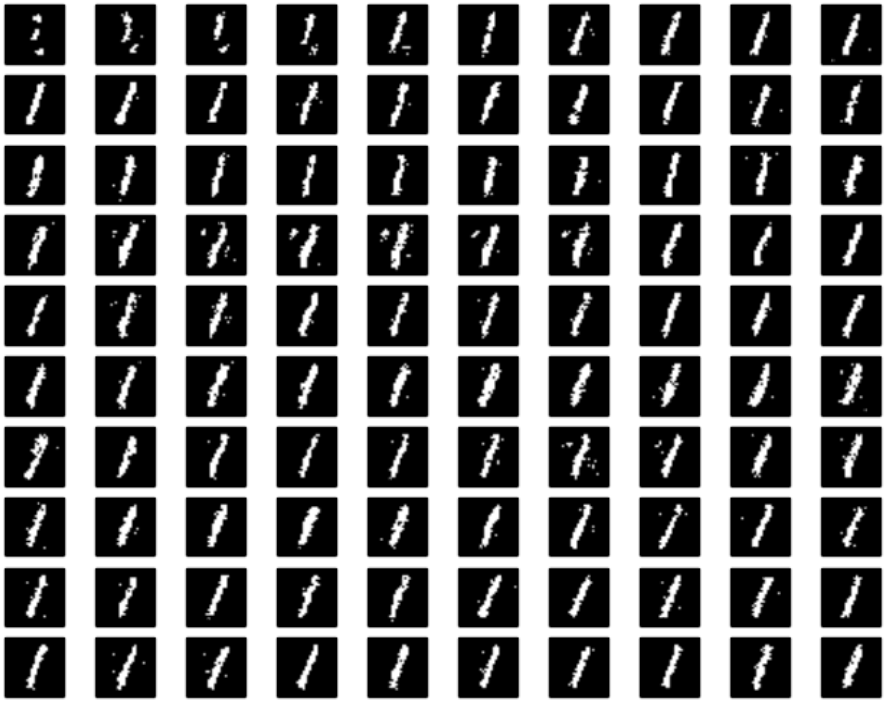
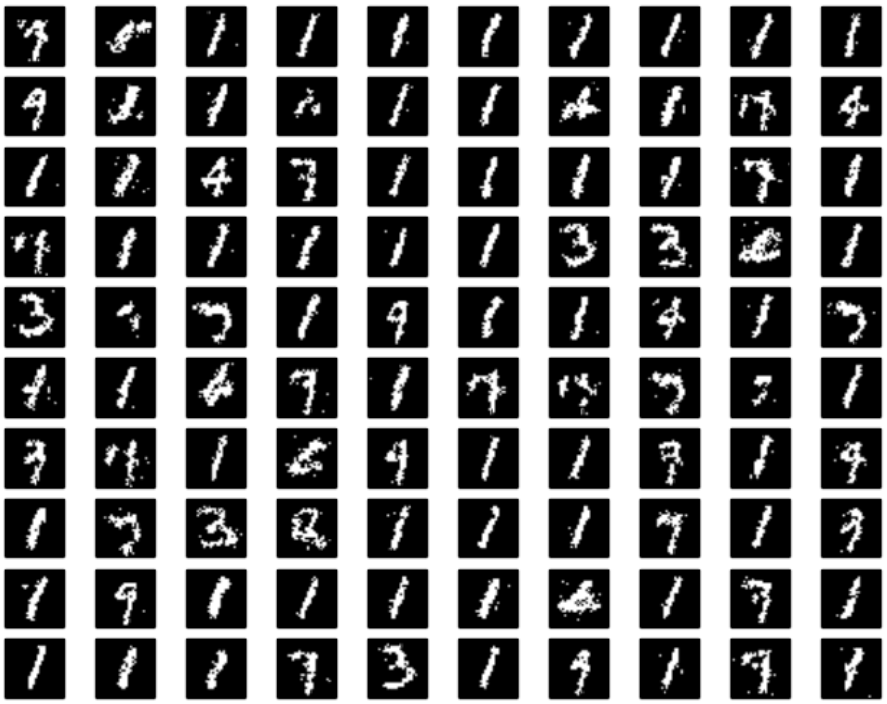
Question 2.

(a) During 500 iterations for a single random initialization we see that successive images try to converge onto a particular digit. Towards the last iteration, we can see a particular well defined digit.



There won't be any mixing of digits for a single random initialization. In next part we see that RBM mixing due to different initializations.

(b) Different initialized values of random matrix gives us different digits in the last sample. Some of the chains converge onto a same digits but each chain has possibility of converging onto any digit to begin with, depending on its initialization value.

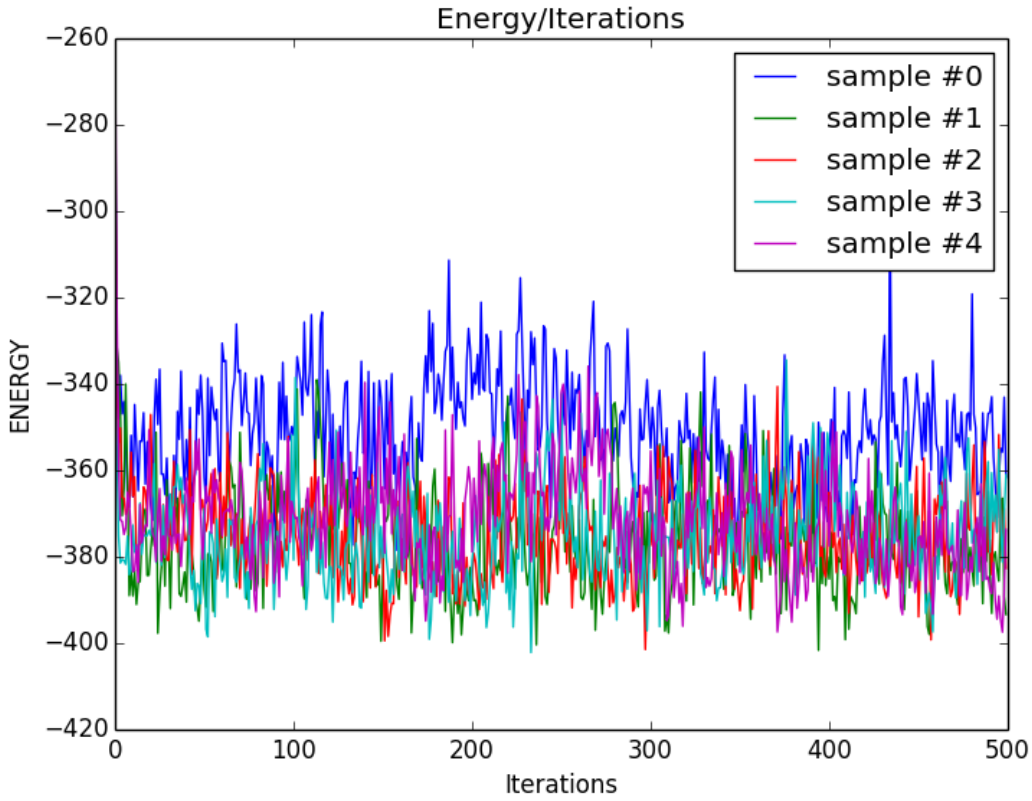


(c) Energy corresponding to different chains try to converge onto some particular energy value. However all of them don't converge at similar values.

We have already seen in part(a) and (b) that for a particular chain, RBM converge onto a single digit and doesn't move between digits. Similarly in energy graph below we can see that for five chains, we have separate energy regions. Each of these energy regions represents a particular digit.

This plot makes us conclude that burn in time for each of the chain is low i.e. chains burn in quickly. It takes nearly 500 iterations for a chain to converge onto a particular energy value giving reasonable visual results for a digit.

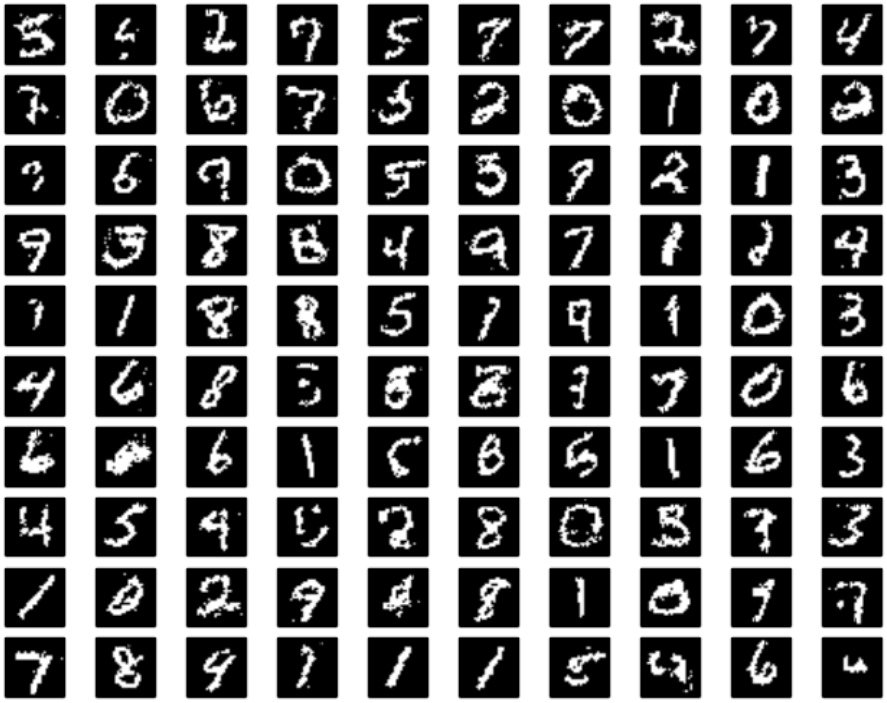
As we observed that for a particular chain, it converges to a single digit having different energy region, each of the consecutive sample drawn from a given chain is highly related to each other. Therefore auto-correlation is high.


Energy/Iterations
sample #0
sample #1
sample #2
sample #3
sample #4
ENERGY
Iterations
−260
−280
−300
−320
−340
−360
−380
−400
−420
0
100
200
300
400
500

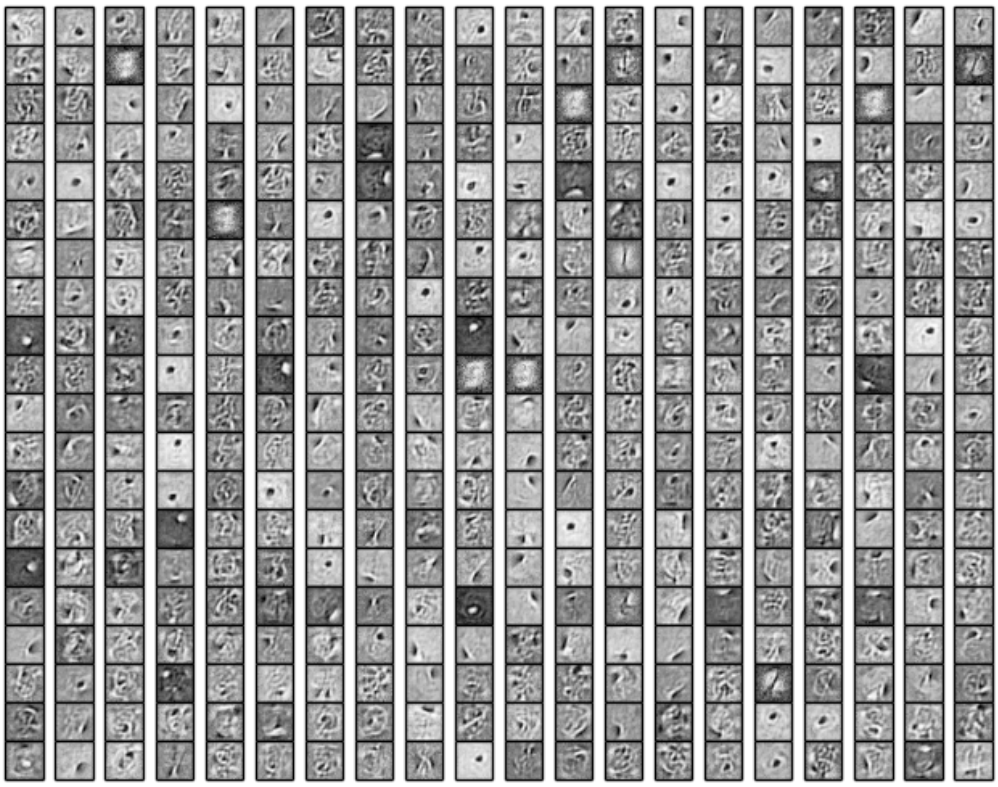
Question 3.

(a) For the given image below, most of the digits are visually recognizable quite easily. However some of the digits are not formed correctly and doesn' match the training images in Figure 1.

The below image is visually better than image shown in Question 2 for K=100. Digits are much better formed with curves and edges quite close to original training image digits. However very few images are not recognizable.



(b)



Question 4.

(a)

I have reported two error rates corresponding to Models with 400 and 100 hidden units:

For 100 hidden units (given model in assignment) : 7.12%
For 400 hidden units (trained in question 3): 3.93%

(b) Error rate with raw pixels as feature: 7.88%

As we can see, lower dimensional features used in part (a) works better than raw pixel values.