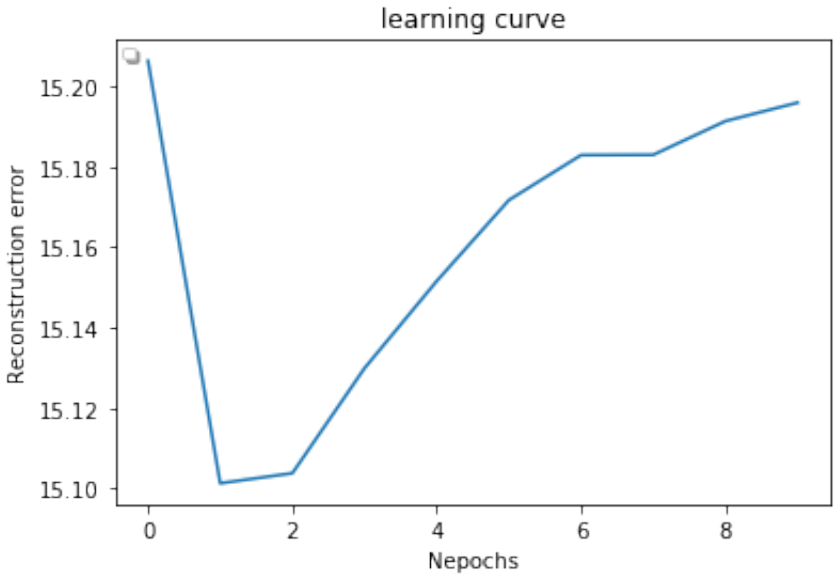
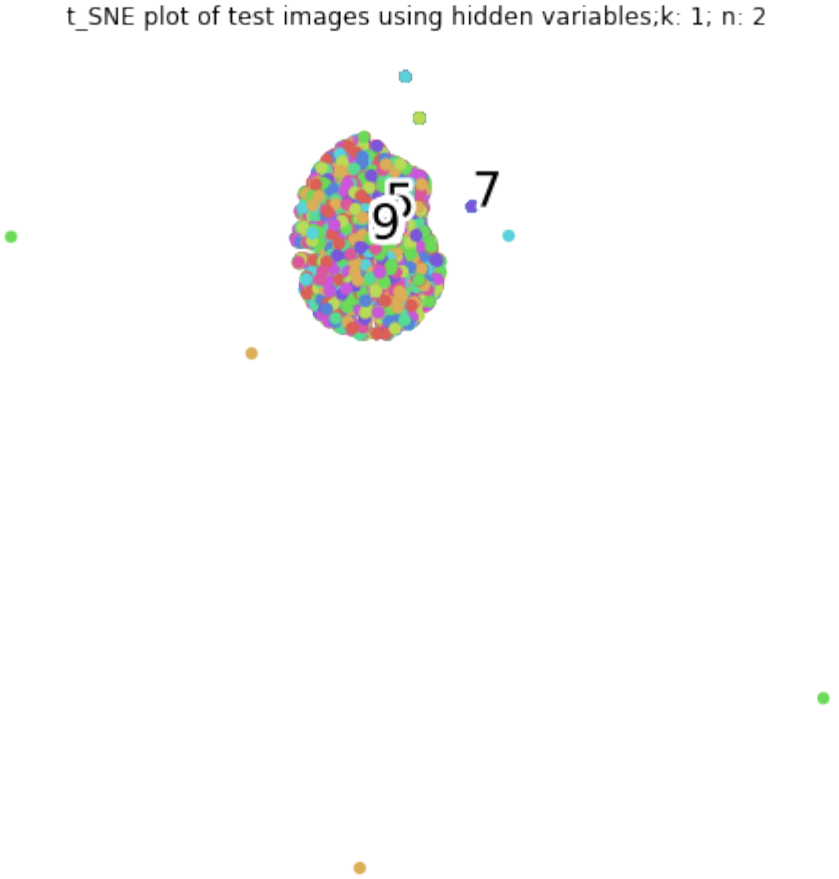
1. RBM

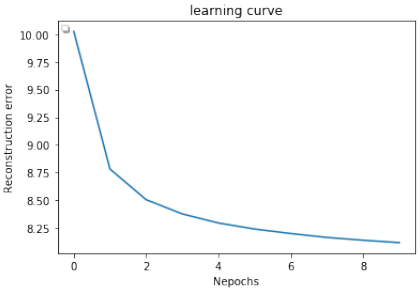
   (a) Use t-SNE to plot the learned representations in a 2-dimensional space (t-SNE will essentially take the n-dimensional representation and plot it in a 2d space such the images which are close in the n-dimensional space will be close in the 2d space also). While plotting use a different color for each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

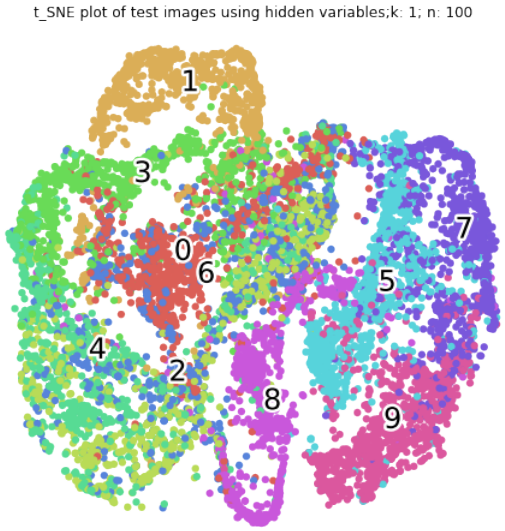
Solution:





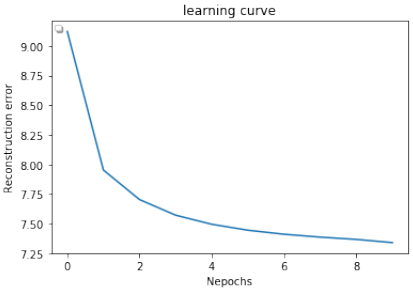
Parameter setting n = 2
Figure 1: The average energy of RBM decreases. Figure 2: t-SNE of learned representations
The Figure 1 curve does not converge. It shows that the RBM with only 2 hidden units is not capable to learn abstraction of given model data.

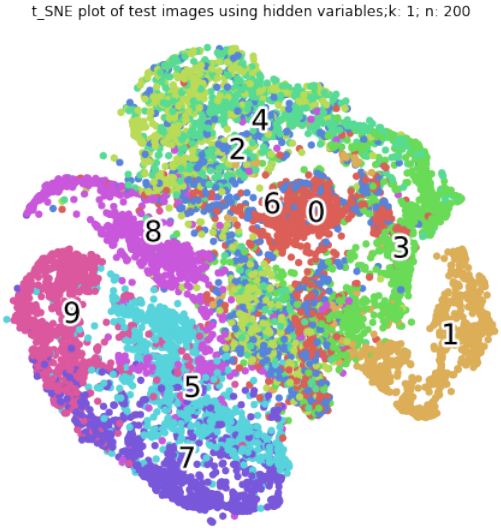




Parameter setting n = 100

Figure 1: The average energy of RBM decreases. Figure 2: t-SNE of learned representations The Figure 1 curve seems not converged but it has dropped much slower after the third epoch. The model has learnt an abstract representation of those images with 100 hidden units.

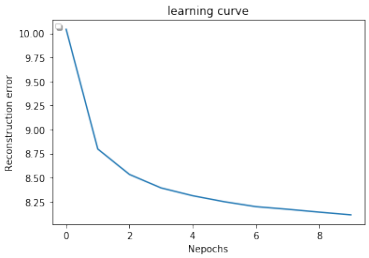




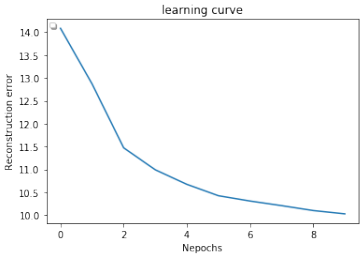
Parameter setting n = 200. The model has learnt good abstract representation of those images with 200 hidden units.Thus as n increases better abstraction is achieved.

(b) 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.

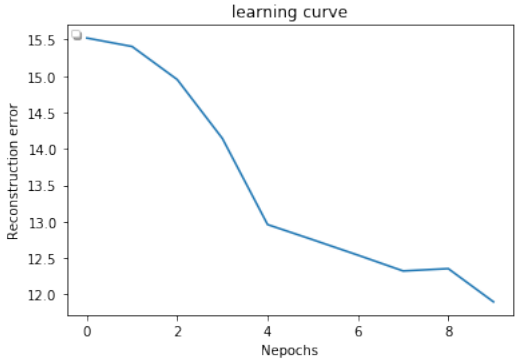
Solution:

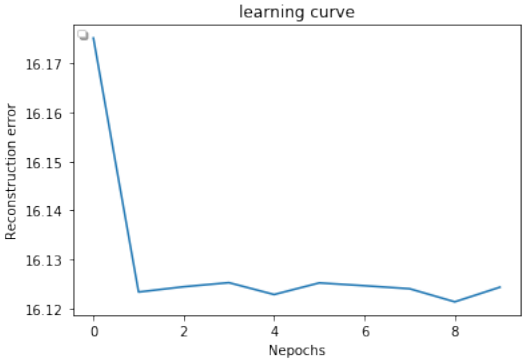


k value 1



k value 3

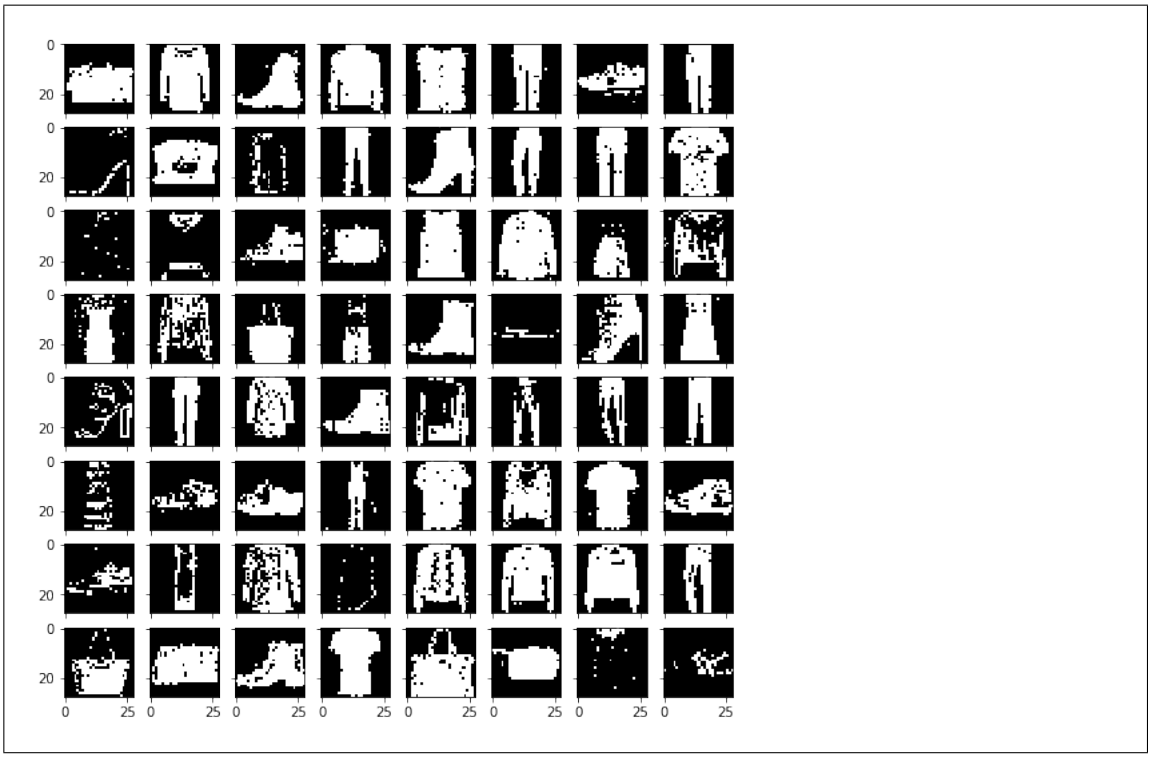




k value 5 and k value 20
As the value of k increases time taken to train increases.

(c) Suppose CD takes m iterations of SGD to converge. Plot the samples enerated by Gibbs chain after every m/64 steps of SGD. Use an 8 x 8 grid to plot these 64 samples.

Solution:



(d)