In this assignment you will implement Restricted Boltzmann machines (RBMs) using only python and numpy. You are not allowed to use tensorflow, theano or any package which supports automatic differentiation (you need to write the backpropagation code yourself).

RBMs can be used to learn hidden representations (h) from the raw features (V). Your task is to train RBMs using the Contrastive Divergence (CD) algorithm. Specifically, given the 784 dimensional (V) binary fashion-MNIST data you need to learn a n-dimensional hidden representation (h). You need to convert the real valued fashion-MNIST data into binary data by using a threshold of 127 (any pixel having a value less than 127 will be treated as 0 and any pixel having a value greater than or equal to 127 will be treated as 1). The specific tasks are listed below:

1. You will use the training data (60000 images) for training the RBM.
2. After training, you will compute the hidden representations for the 10000 test images.
3. You need to submit your code and accompanying scripts and a report containing the plots mentioned below.

Evaluation Criteria

1. **5 Marks** 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
2. **5 Marks** 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.
3. **5 Marks** Suppose CD takes m iterations of SGD to converge. 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.
4. **Extra Credits: 10 Marks** 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 ?