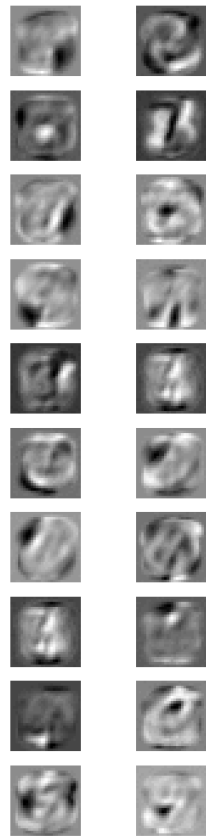


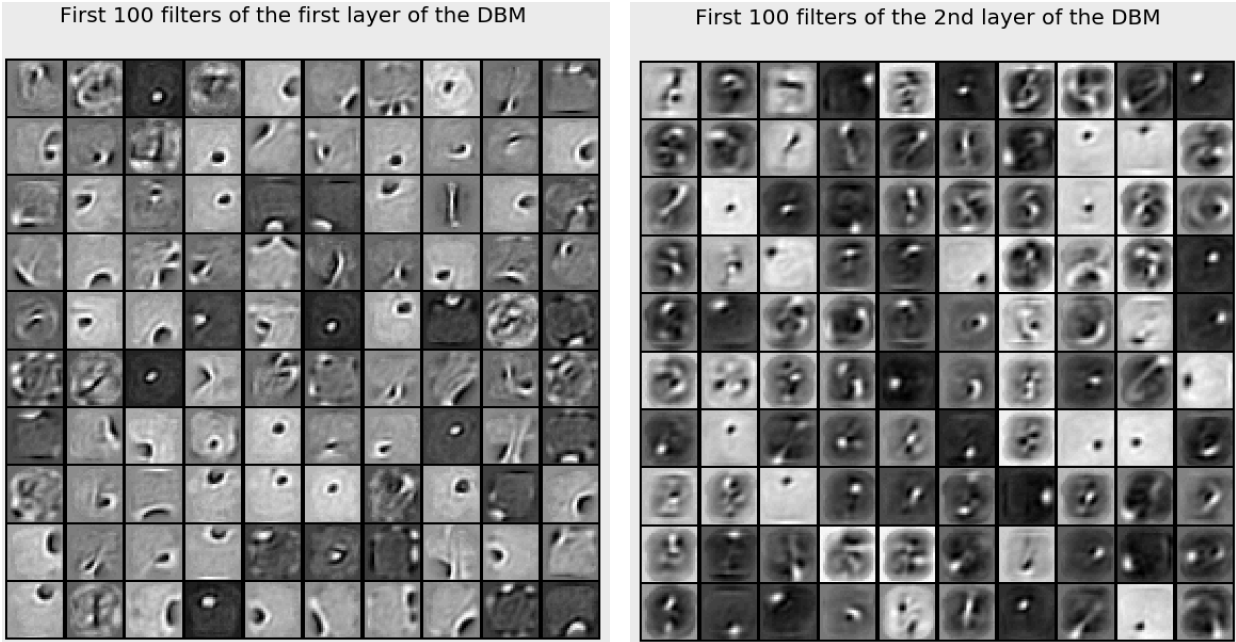
1 Train a DBM on the MNIST database.
Show the filters from the previously trained RBM and the DBM, what is the difference?
Can you explain the difference between filters of the first and second layer of the DBM?

How is DBM trained ?
Boltzmann machine uses randomly initialized Markov chains to approximate the gradient of the likelihood function which is too slow to be practical.
DBM uses greedy layer by layer pre training to speed up learning the weights. It relies on learning stacks of Restricted Boltzmann Machine with a small modification using contrastive divergence.
The key intuition for greedy layer wise training for DBM is that we double the input for the lower-level RBM and the top level RBM.
Lower level RBM inputs are doubled to compensate for the lack of top-down input into first hidden layer.Similarly for top-level RBM, we double the hidden units to compensate for the lack of bottom-up input.
For the intermediate layers, the RBM weights are simply doubled.

filters of RBM



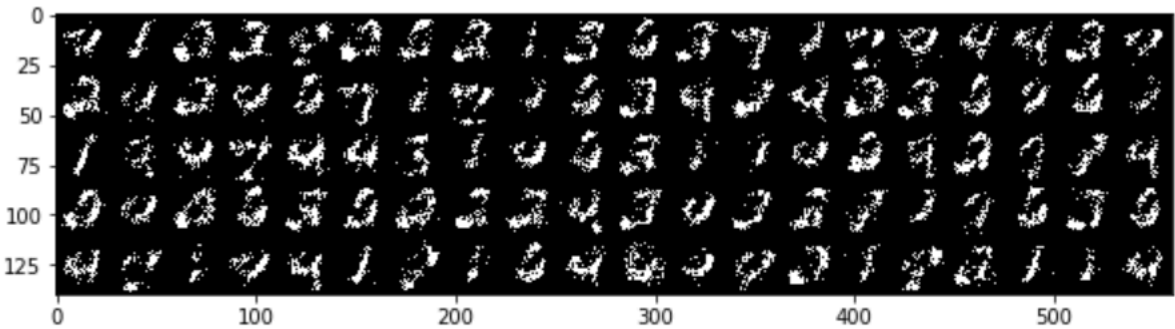
filters of DBM



Difference between RBM and DBM filters: DBM is built upon RBM to **increase its representation power by increasing depth.**

Difference between filters of the 1st and 2nd layer of the DBM :filters on the second layers give more details

2 Sample new images from the DBM. Is the quality better than the RBM from the previous exercise and explain why?



Samples generated by DBM after 100 Gibbs steps

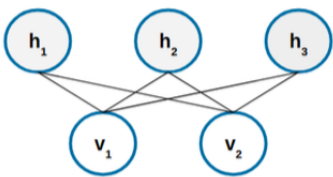


Deep belief network is a deep architecture built upon RBM to **increase its representation power by increasing depth**.

In a DBN, two adjacent layers are connected in the same way as in RBM. The network is trained in a greedy, layer-by-layer manner, where the bottom layer is trained alone as an RBM, and then fixed to train the next layer.

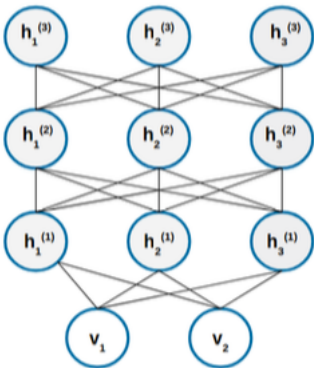
After all layers are pre-trained, the resulted network contains one RBM at the top layer while the rest layers form a directed neural network. There is no joint training for all layers in DBN.

To generate images, we need to first run Gibbs sampling in the top RBM till convergence and then propagate it down to the bottom layer. DBN has also been extended to use convolutional RBMs for better scalability and higher quality features.



RBM has binary-valued hidden and visible units, and consists of a matrix of weights $W = (w_{i,j}) \in \mathbb{R}^{m \times n}$ associated with the connection between hidden unit h_j and visible unit v_i , as well as bias weights (offsets) a_i for the visible units and b_j for the hidden units.

The term restricted comes from the fact that there no hidden-to-hidden and visible-to- visible connections.



Deep Boltzmann machines can be understood as a series of restricted Boltzmann machines stacked on top of each other [3]. The hidden units are grouped into a hierarchy of layers, such that there is full connectivity between subsequent layers, but no connectivity within layers or between non-neighbouring layers.