# Chapter-28

Sunday, March 21, 2021 11:50 AM

## Deep Generative Models.

### Deep Directed Networks-

If all nodes are binary and all CPDs are logistic functions, this is called a sigmoid belief net.

$$p(h_1,h_2,h_3,v|\Theta) = \prod_i Ber(v_i|sigm(h_1^T w_{0i}))\prod_j Ber(h_{1j}|sigm(h_2^T w_{1j})).$$

$$\prod_k Ber(h_{2k}|sigm(h_3^T w_{2k}))\prod_l Ber(h_{3l}|w_{3l}).$$

Inference is intractable since the posterior on hidden nodes is correlated due to the explaining away.

### Deep Boltzmann Machines-

Construct a deep undirected model by stacking a series of RBMs on top of each other.

$$p(h_1,h_2,h_3,v|\Theta) = \frac{1}{Z(\Theta)}\exp\Big(\underbrace{\sum_{ij} v_i h_{1j} W_{1ij}}_{\text{visible-hidden interactions}} + \underbrace{\sum_{jk} h_{1j} h_{2j} W_{2jk} + \sum_{kl} h_{2k} h_{3l} W_{3kl}}_{\text{hidden-hidden interactions}}\Big).$$

One can efficiently perform block Gibbs sampling since all nodes in each layer are conditionally independent of each other given the layers above and below.

### Deep Belief Networks-

"Partially directed and partially undirected"

$$p(h_1,h_2,h_3,v|\Theta) = \prod_i Ber(v_i|sigm(h_1^T w_{1i}))\prod_j Ber(h_{1j}|sigm(h_2^T w_{2j}))$$

$$\frac{1}{Z(\Theta)}\exp\Big(\sum_{kl} h_{2k} h_{3l} W_{3kl}\Big).$$

For a model of the form $p(h_1,h_2,v|W_1)$,

$$p(h_1,v|W_1) = \sum_{h_2} p(h_1,h_2,v|W_1).$$

$$\Rightarrow p(h_1,v|W_1) = \frac{1}{Z(W_1)}\exp(v^T W_1 h_1).$$

This is equivalent to an RBM.

Thus, we can infer the posterior $p(h_1|v,W_1)$ in the DBN exactly as in the RBM.

Now, $p(h_1|W_1)$ is the complementary prior.

$$\Rightarrow p(h_1,v|W_1) = p(v|h_1)\,p(h_1|W_1)$$

The top-down inference in a DBN is not tractable, so DBNs are usually used in a feedforward manner.

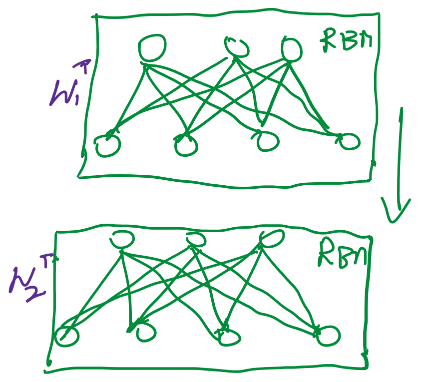
### Greedy layer-wise learning of DBNs -

1) Fit an RBM to learn $W_1$.

2) Unroll the RBM into a DBN with 2 hidden layers.

3) Freeze $W_1$ and let $W_2$ be untied. Learn $p(h_1|W_2)$ by fitting a second RBM with inputs as activation of hidden units $E[h_1|v,W_1]$.

4) Continue to add more layers.




Following greedy layer-wise training strategy, it is standard to "fine tune" the weights using a technique called backfitting.