Introduction to Neural Networks

Johns Hopkins University
Engineering for Professionals Program
605-447/625-438
Dr. Mark Fleischer

Module 11.1: Restricted Boltzmann Machines

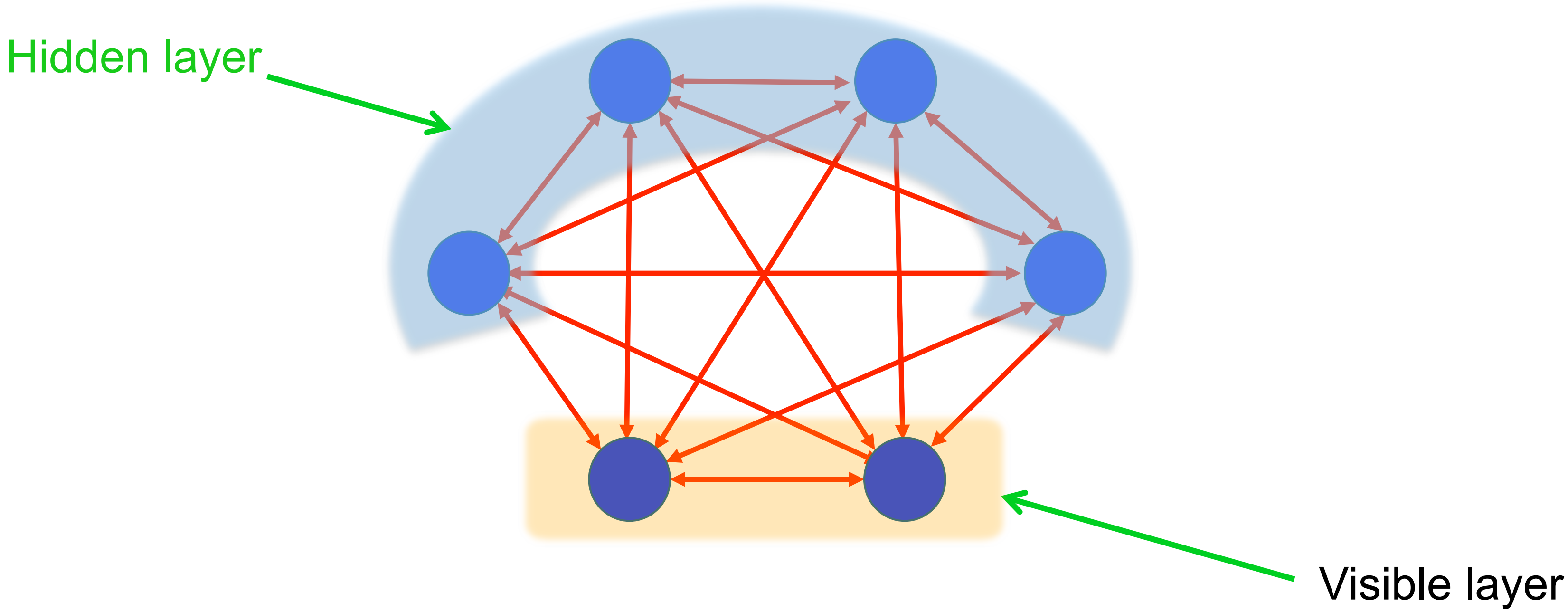
What We've Covered So Far

- Boltzmann machines are essentially stochastic versions of Hopfield networks.
- Stochastic assignment of node states using a binary value.
- Associate a performance metric via the energy or consensus function with network configurations.
- Capability to 'anneal' a BM to solve various optimization problems.

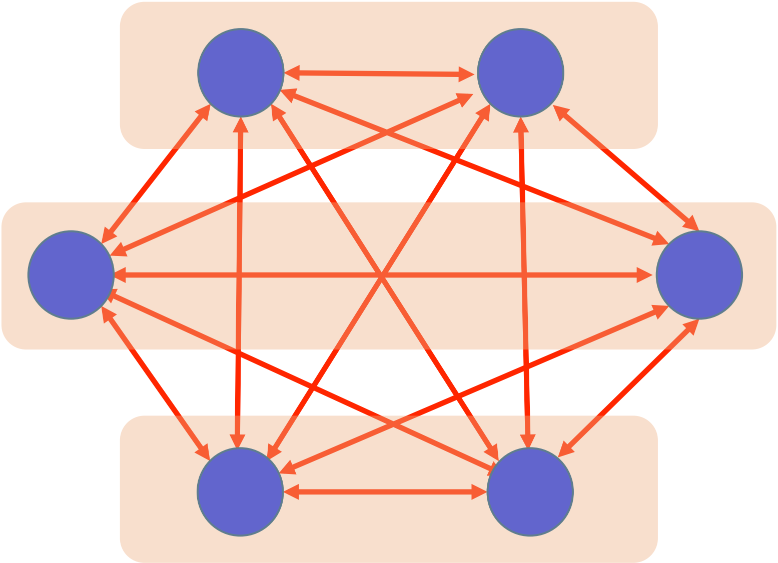
Model Relationships
for Recurrent Networks

Deterministic		Stochastic
Hopfield Networks	→	Boltzmann Machines
Binary Associative Memories	→	Restricted Boltzmann Machines

Restricted Boltzmann Machines



Restricted Boltzmann Machines

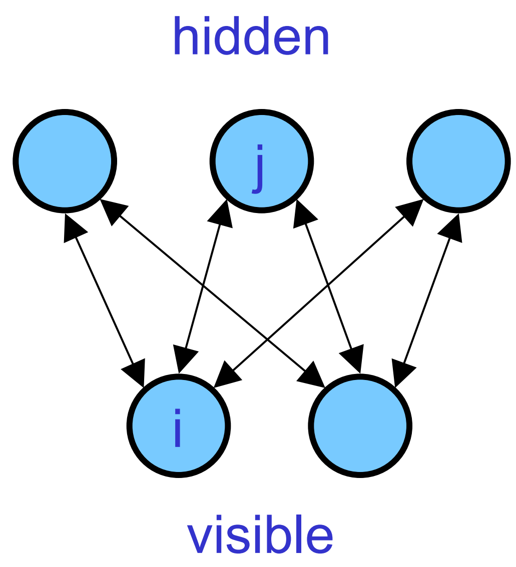


Sort of a multi-layer BAM!

Configured as a *Belief Network*

Restricted Boltzmann Machines

- We restrict the connectivity to make learning easier.
  - Only one layer of hidden units.
    - We will deal with more layers later
  - No connections between hidden units.
- In an RBM, the hidden units are conditionally independent given the visible states.
  - So we can quickly get an unbiased sample from the posterior distribution when given a data-vector.
  - This is a big advantage over directed belief nets



From Hinton 2007

Purpose of RBMs

- Many possible applications.
- Essentially attempts to establish a useful association between visible vectors and hidden vectors similar in nature to how BAMs function.
  - used in factor analysis
  - character recognition
  - many others---active area of research
- Stochastic states allow for greater flexibility and noise

What Do RBMs Do?

- Just like BAMs, they attempt to 'reconstruct' a training vector.
  - In BAMs these were the 'A' vectors or the 'data'.
  - Noisy or variable training vectors probabilistically produce feature detectors in the hidden layer which then probabilistically produce 'reconstructions' of the training vectors.

Binary Associative Memories

$$A_1 \longleftrightarrow B_1$$

$$A_2 \longleftrightarrow B_2$$

$$\vdots \qquad\qquad \vdots$$

$$A_n \longleftrightarrow B_n$$

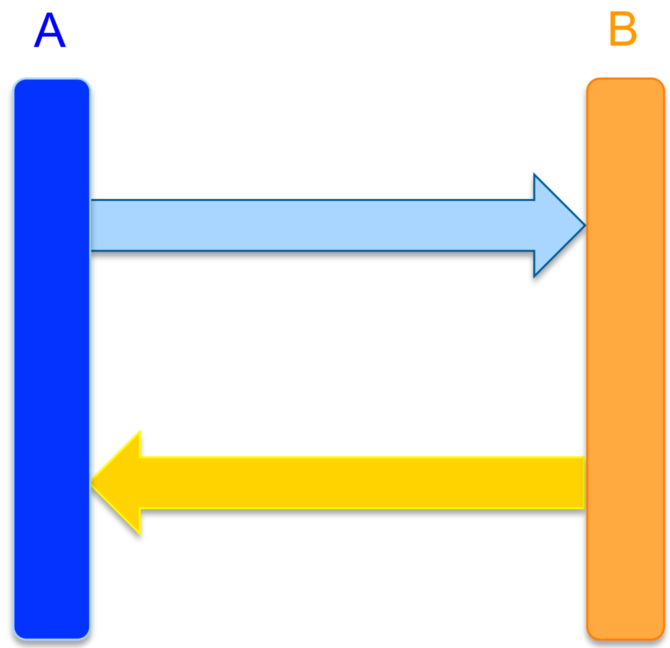
Goal: Noisy A_1 produces a correct B_1 which then produces a correct A_1.

$$\widetilde{A}_1 \rightarrow B_1 \rightarrow A_1$$

Binary Associative Memories

- Present a noisy A as input to the A nodes.
- The A nodes produce outputs and are presented to the B nodes.
- The B nodes produce outputs and are presented back to the A nodes.

Binary Associative Memories
Restricted Boltzmann Machines

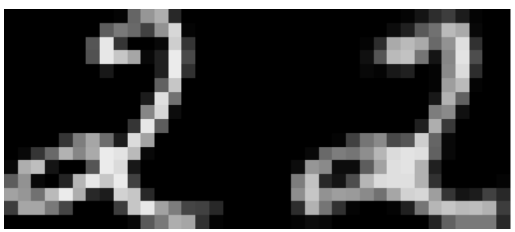


Only Stochastically!

How well can we reconstruct the digit images from the binary feature activations?

Data

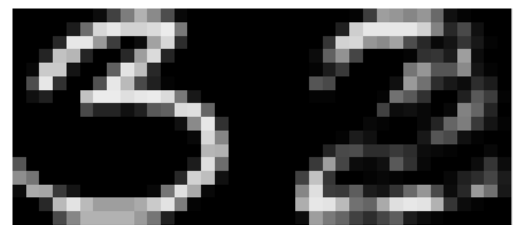
Reconstruction from activated binary features



New test images from the digit class that the model was trained on

Data

Reconstruction from activated binary features



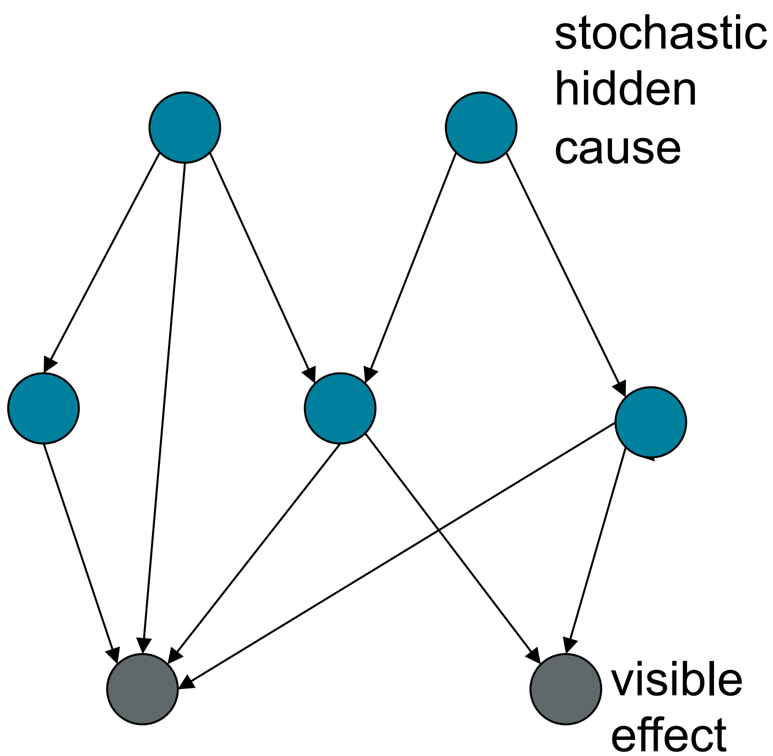
Images from an unfamiliar digit class
(the network tries to see every image as a 2)

From Hinton 2007

Belief Nets

- A belief net is a directed acyclic graph composed of stochastic variables.
- We get to observe some of the variables and we would like to solve two problems:
- The inference problem: Infer the states of the unobserved variables.
- The learning problem: Adjust the interactions between variables to make the network more likely to generate the observed data.

From Hinton 2007

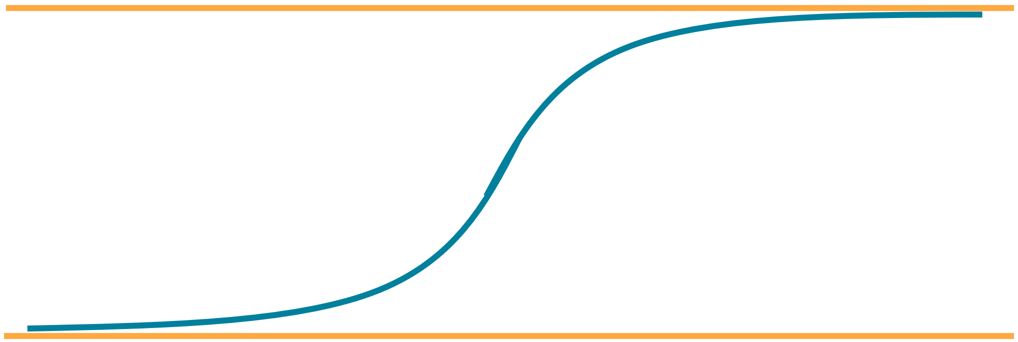


We will use nets composed of layers of stochastic binary variables with weighted connections

Dynamics

Recall the Sigmoid activation function

$$\frac{1}{1+e^{-S_i/T}}$$



Again, we're dealing with stochastic neurons.
What does this curve remind you of?

Dynamics

Set cell activation (state) according to:

$$x_i = \begin{cases} 1 & \text{with probability } p_i = \dfrac{1}{1+e^{-S_i/T}} \\ 0 & \text{with probability } 1 - p_i \end{cases}$$

The Energy/Consensus Function

Define the energy function *E*. With the BM we used

$$E = -\sum_{i<j} w_{ij} x_i x_j - \sum_i \theta_i x_i - \sum_j \xi_j x_j$$

In RBMs, only the connections between layers is important.

$$E = -\sum_{i,j} w_{ij} v_i h_j - \sum_i \theta_i v_i - \sum_j \xi_i h_i$$

The Energy of a joint configuration

(ignoring terms with biases)

binary state of visible unit i

binary state of hidden unit j

$$E(\mathbf{v},\mathbf{h}) = -\sum_{i,j} v_i h_j w_{ij}$$

Energy with configuration v on the visible units and h on the hidden units

weight between units i and j

$$\frac{\partial E(\mathbf{v},\mathbf{h})}{\partial w_{ij}} = -v_i h_j$$

From Hinton 2007 (modified)

Think in terms of consensus

$$C(\mathbf{v},\mathbf{h}) = \sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij} v_i h_j$$

If we want to strengthen an association where both v_i and h_j are 1s,

make w_{ij} a large positive number.

If we want to strengthen an association where v_i and h_j have opposite states,

make w_{ij} a large negative number.

Summary

- RBMs are stochastic versions of BAMs.
  - Two layers: *visible* and *hidden*
- Use probabilistic machinery of Boltzmann machines.
  - Probability of a node's state is based on sigmoid function with activity value based on weighted inputs from 'the other set of nodes'
  - Energy of a (v, h) pair is defined as sum of the weighted products of visible node states and hidden node states.