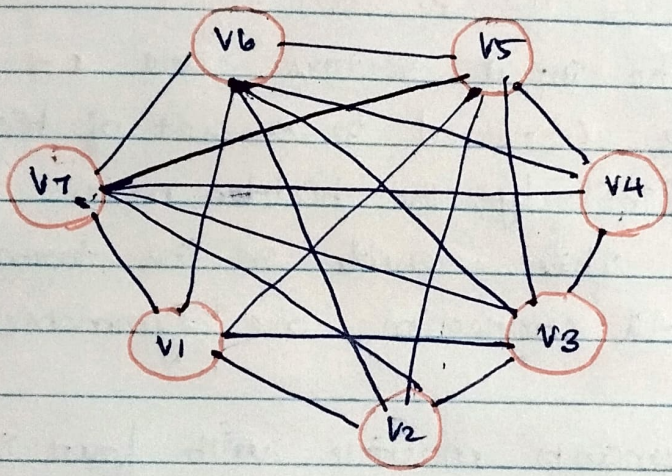
Boltzmann Machine.

* Boltzmann machines are symmetrically connected recurrent neural network
* They are said to symmetrically connected because they are undirected edges between the nodes
* The weight from node i to node j is the same as the weight from node j to node i.
* The input vectors in Boltzmann machine is binary vectors.



Structure of a Boltzmann machine.

Relation to Hopfield Network.

i) The similarity between Hopfield networkes and Boltznmann machines is that

* they both have a fully connected network of nodes
* energy is associated with the entire network

ii) Boltzmann machines are stochastic i.e. output of boltzmann machine is a probability distribution and

not continuous values as in regression or categorical values as in classification.

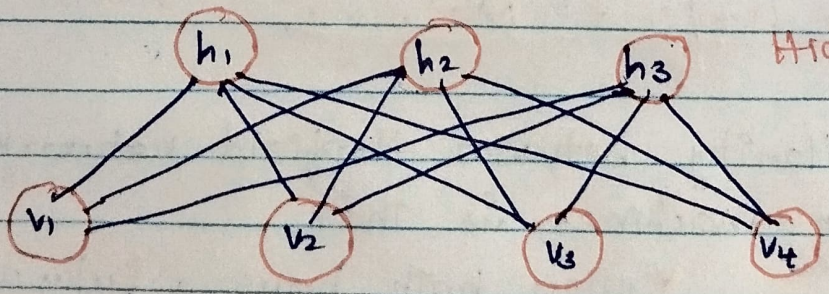
RBM Architecture.

RBM helps to perform unsupervised training of data i.e it is possible to train an RBM on an unlabeled training data.

* RBM is a special case of Boltzmann machine which has two layers
  i) visible (input) layer.
  ii) hidden layer
* The input vectors are not connected to each other and are connected to a set of hidden nodes
* Similarly, the hidden nodes are not connected to one another. This results in the term restricted
* All connection are symmetric

Example:

Restricted Boltzmann machine with four input nodes (v_1, v_2, v_3, v_4) and three hidden nodes (h_1, h_2, h_3)



This bipartite graph because it has two set of nodes, where one set of nodes is fully connected to the nodes in the second set.

Energy-Based Model:

* Boltzman machines are closely associated with quantum physics.
* Energy is associated with the Restricted Boltzmann machines
* The state of the entire network is now given in terms of the energy of the network, termed the energy function
* Energy is determined by configuration of variables.
* The variables associated with the RBM are the
  i) values in visible node
  ii) value in the hidden node
* The nature of energy function is that energy becomes large, when the compatibility between the variable is low.
* The objective is to increase the compatibility between the variables.
* Energy value is decreased iteratively till it reaches the local minima

Gibbs Distribution:

* The output of an RBM is also called Gibbs distribution.
* The probability distribution with which a system given by x in a certain state as

$$P(x) = \frac{e^{-E(x)}}{Z}$$

* Z is the normalizing factor and is called the partition function. It is given by

$$Z = \sum_{x} e^{-E(x)}$$

* In RBM, the input are the values on the visible (v) and hidden neurons (h)
* The join probability for the state given by v and h is .

$$P(v,h) = \frac{e^{-E(v,h)}}{Z}$$

where

$$Z = \sum_{x} e^{-E(v,h)}$$

* v is input vector
* h is the hidden layer vector.

* We conider all possible pairs between the visible and hidden units to calculate Z
* The joint probability P(v,h) is difficult to compute because the number of possible pairs between v and h would be vxh
* Therefore Gibbs sampler is used to find the probability.