The Boltzmann Machine

Olivia Guest

February 20, 2014

1 Basic Concepts

A BM is a type of stochastic constraint satisfaction neural network based on statistical thermodynamics (Ackley, Hinton, & Sejnowski, 1985). It is an extension of the Hopfield neural network model that incorporates simulated annealing, and is composed of units connected using bidirectional symmetric links (Hinton & Sejnowski, 1986). Units have a probabilistically determined binary state that is calculated based on its neighbours' states and the weights on the links to it. A connection weight can take any positive or negative real value.

The units are divided into visible, v, and hidden, h. In the most general case, all the units are fully connected, meaning every unit is connected to every other unit except itself; giving a total of $\frac{N \times N - N}{2}$ weights, where N is the number of units. Visible units function as the interface between the BM and its environment. Hidden units capture "underlying constraints in the ensemble of input vectors that cannot be represented by pairwise constraints among the visible units" (Ackley et al., 1985, p. 154) and are always free, i.e., unclamped. After training, a non-empty subset of the visible units may be clamped, in which case the network's weights will calculate the remaining free unit states, resulting in pattern completion (Hinton & Sejnowski, 1986).

2 Interpretation of Network States

The activity of a BM unit can be interpreted as acceptance of a domain hypothesis, if it is on, or rejection, if it is off. The weight of a link between two units represents a weak constraint on the two hypotheses; a positive weight means the two hypotheses support each other (i.e., if one unit fires, it is likely the other also will) and a negative weight indicates neither hypothesis should be accepted (Hinton, Sejnowski, & Ackley, 1984).

When the visible units are clamped the BM seeks to locate a configuration of hypotheses that minimises the global energy. This energy, the extent to which the BM's configuration violates the constraints imposed by the environment (Hinton et al., 1984), is defined as:

$$E = -\sum_{i<j} w_{ij} s_i s_j + \sum_i \theta_i s_i \tag{1}$$

where w_{ij} is the weight, or connection strength, between unit i and j, s_i is the state of unit i (on is 1 and off is 0), and θ_i, the threshold[1], is a self-regulatory parameter applied to each unit (Ackley et al., 1985).

Finding the global minimum in the energy landscape is not guaranteed. At best what can be done is to perform a search for local energy minima, by switching every unit's state to the one which lowers most the total energy (Hinton & Sejnowski, 1986). Due to the network's nature this energy gap, between accepting the kth hypothesis and rejecting it, can be calculated locally, on per unit basis, using:

$$\Delta E_k = \sum_i w_{ki} s_i - \theta_k \tag{2}$$

where k is a unit that will fire if its input is above its threshold, θ_k (Ackley et al., 1985).

[1] Also known as a bias, b_i, if it is inverted: $b_i = -\theta_i$.

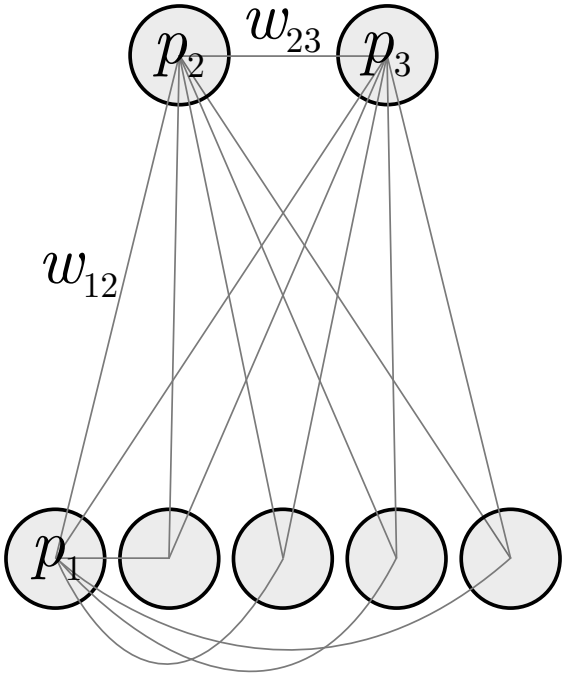

Figure 1: A BM with the connections of units 1—3 depicted only, with only two weights explicitly labelled. The probabilities, p_i, used to calculate the states of these three units are displayed, however the two layers' states should be further labelled as v_i and h_j.

3 Simulated Annealing, Thermal Equilibrium, and Learning

Local minima are by definition not optimal and thus, the BM must be able to avoid or escape from them. This is accomplished by allowing units to fire probabilistically; amounting to introducing noise into the network, permitting jumps between states of different energy (Hinton et al., 1984). The probability that a unit k will fire is set to:

$$p_k = \frac{1}{1 + e^{-\Delta E_k / T}} \tag{3}$$

where ΔE_k is k's energy gap (see Equation 2), and T is the temperature (Ackley et al., 1985). As T approaches 0, thermal equilibrium, the relative probability of two global states[2] follows the Boltzmann distribution. Due to the inherent properties of this distribution, in order to reach equilibrium a process called simulated annealing must be applied to the network[3].

The equation used to update the value of each weight, the learning rule[4] is:

$$\Delta w_{ij} = \epsilon(p_{ij} - p'_{ij}) \tag{4}$$

where ϵ, the learning rate, moderates the effect the difference between the two probabilities has on w_{ij} (Ackley et al., 1985), p_{ij} is the probability of units i and j both being on when all the visible units are clamped, the plus phase, and p'_{ij} is the probability of the units co-occurring when the BM is running freely at thermal equilibrium, the minus phase (Hinton et al., 1984).

The learning rule, as mentioned, uses only locally available information to optimise a global energy measure (Hinton et al., 1984). It has the ability to make the spontaneous behaviour of the network in the minus phase mimic that of the plus phase, by only observing how often units i and j are on at thermal equilibrium (Hinton & Sejnowski, 1986).

[2] A global state is a set of all unit states; there are 2^{v+h} possible global states.

[3] For a detailed account of the reasons behind this, see Hinton and Sejnowski (1986).

[4] For the mathematical derivation of the BM learning algorithm see the appendix in Hinton and Sejnowski (1986) or in Ackley et al. (1985).

References

Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A Learning Algorithm for Boltzmann Machines. *Cognitive Science*, *9*, 147–169.

Hinton, G., & Sejnowski, T. (1986). Learning and relearning in boltzmann machines. *MIT Press, Cambridge, Mass.*, *1*, 282–317.

Hinton, G., Sejnowski, T., & Ackley, D. (1984). *Boltzmann machines: Constraint satisfaction networks that learn* (Tech. Rep.).