Network Inference for Directed Graphs with Stochastic Hidden Units

John Hertz, Nordita, Stockholm University and Royal Institute of Technology
Joanna Tyrcha, Mathematical Statistics, Stockholm University
jhertz@nordita.org, joanna@math.su.se

A good deal of recent work has dealt with the problem of inferring the connections in a network from a record of the activity of its units. In particular, attention has focused on statistical modeling of neural networks based on recorded spike trains. However, this work has assumed that data are available from all the units in the network, a condition that never applies in actual experiments. Here we lay a foundation for treating more realistic situations by developing learning algorithms for networks of binary stochastic units containing "hidden" in addition to "visible" units. For Gibbs equilibrium networks, with symmetric connection matrices, this problem is solved by the classic Boltzmann learning algorithm. For the more general case of asymmetric connection matrices, this problem was also solved many years ago ("back-propagation in time") for networks of continuous-valued units. However, learning in networks of stochastic binary units with arbitrary connection matrices has not been treated previously. We derive an algorithm for this case here.

We consider networks of units can take on the values ± 1 and are updated synchronously according to the stochastic rule $P[S_i(t+1)|\mathbf{S}(t)] = \frac{1}{2}[1 + S_i(t+1)\tanh H_i(t)]$, with $H_i(t) = \sum_j J_{ij} S_j(t)$. Denoting the N_v visible units by s_i and the N_h hidden ones by σ_i, the likelihood of a history of the visible units s is obtained by marginalizing out the hidden units, and the learning rules are found by gradient ascent on the resulting log likelihood $\log P[\mathbf{s}] = \log \sum_{\boldsymbol{\sigma}} P[\mathbf{s}, \boldsymbol{\sigma}]$. They have the form

$$\Delta J_{ij} \propto \frac{\partial \log P[\mathbf{s}]}{\partial J_{ij}} = \sum_t \sum_{\boldsymbol{\sigma}} [S_i(t+1) - \tanh H_i(t)] S_j(t) P[\boldsymbol{\sigma}|\mathbf{s}]. \tag{1}$$

The weight $P[\boldsymbol{\sigma}|\mathbf{s}] = P[\mathbf{s}, \boldsymbol{\sigma}]/P[\mathbf{s}]$ in this average is like a Boltzmann weight for a statistical-mechanical problem for the hidden units with an energy $E_{\mathbf{s}}[\boldsymbol{\sigma}] = -\log P[\mathbf{s}, \boldsymbol{\sigma}]$. If there are no hidden-to-hidden connections, $P[\mathbf{s}, \boldsymbol{\sigma}]$ factorizes into a product of independent factors for each time step in the data, and the statistical-mechanical problem for each time step has N_h variables. In the general case, the problem has $N_h T$ variables, where T is the data length. The energy $E_{\mathbf{s}}[\boldsymbol{\sigma}]$ contains terms describing "external fields" acting on each $\sigma_i(t)$ from the visible data $s_j(t \pm 1)$ at both the previous and subsequent time steps, terms which couple each $\sigma_i(t)$ with all $\sigma_j(t \pm 1)$, and nonlinear couplings among the different σ_i at each t. Using these Boltzmann weights to evaluate the averages in (1) constitutes the exact algorithm for finding the couplings.

We also derive mean field equations for the auxiliary statistical-mechanical problem. The effective fields determining the "magnetizations" $m_i = \langle \sigma_i \rangle$ are combinations of forward messages and back-propagated errors:

$$\begin{aligned} \tanh^{-1} m_a(t) &= \sum_j J_{aj} s_j(t-1) + \sum_b J_{ab} m_b(t-1) \\ &+ \sum_j [s_j(t+1) - \tanh H_j(t)] J_{ja} + \sum_b [m_b(t+1) - \tanh H_b(t)] J_{ba} \end{aligned} \tag{2}$$

This is a generalization of an old formulation of back-propagation learning for conventional networks with a single hidden layer in terms of explicit "internal representations".