Learning from mistakes

Dante R. Chialvo*† and Per Bak*‡

‡*Niels Bohr Institute, Blegdamsvej 17, Copenhagen, Denmark.* †*Division of Neural Systems, Memory and Aging, University of Arizona, Tucson, AZ 85724, USA.* **Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA.*

(July 27, 1997)

A simple model of self-organised learning with no classical (Hebbian) reinforcement is presented. Synaptic connections involved in mistakes are depressed. The model operates at a highly adaptive, probably critical, state reached by extremal dynamics similar to that of recent evolution models. Thus, one might think of the mechanism as synaptic Darwinism.

It is widely believed that learning in the brain resides in alterations of synaptic efficacy. Without exception, contemporary formulations of such learning follows Hebb's ideas [1] of reinforcement: synaptic connections among neurons excited during a a given firing pattern are strengthened by a process of long term potentiation (LTP).

However, long term synaptic depression (LTD) in the mammalian brain is almost as prevalent as potentiation, but there appears to be little or no understanding of its functional role. Working hypotheses covers a wide range, where depression is given always an auxiliary function to potentiation [2]. A recent review [3], reflecting the current variety of ideas regarding the functional role of LTD, speculates: "Although it is conceivable that LTP is the critical phenomena used for storing information, and that LTD may exist simply to reset LTP, it must be noted that it is also conceivable for the converse to be true."

We present an alternative to Hebbian learning. Turning things upside down, we suggest that LTD is, in some instances of learning and development, the fundamental mechanism with LTP playing a secondary role. This view is supported by studies of a simple neuronal learning model. There are two fundamental differences between the classical view of learning by reinforcement and the view discussed here:

1) Learning by reinforcing good responses is a process that by definition never stops. There is not an explicit rule that ends the reinforcement whenever the goal has been reached. On the other hand, if learning proceeds only by correcting mistakes it implies a process that stops as soon as the goal is achieved. This prevents formation of "deep holes", i.e. highly stable states from which adaptation to new rules is difficult and slow, requiring, perhaps, a significant amount of random noise.

2) If an adaptive system is placed on a new environment, or otherwise subjected to learn something new, the likelihood of making mistakes is generally larger than the chance to be initially right. Therefore, the opportunity to shape synapses is larger for the adaptive mechanism that only relies on mistakes, leading to faster convergence.

In order to develop these ideas, a model of a adaptive neural structure has been constructed. Although it is only a caricature of a real brain, all the ingredients are biologically reasonable and correspond to well- documented physiological processes. The model is completely self-organised with no need for external computation of synaptic strengths, in contrast to, for instance, feed-forward and back-propagation neural networks. All control mechanisms are local at the post-synaptic site of the active neurons, and information is fed back globally to all neurons.

Each neuron receives input from, and sends output to, several other neurons. Just about any arbitrary architecture, for instance a completely random one, can be chosen. For descriptive purposes, however, consider a two layer network where K represents the outputs, I the inputs and J the middle layer (Figure 1). Each input is connected with each neuron in the middle layer which, in turn, is connected with each output neuron, with weights W representing the synaptic strengths. The network must learn to connect each input with the proper output for any arbitrary associative mapping. The weights are initially randomised, $0 < W < 1$.

In order to achieve efficient self-organised learning, it is essential to keep the activity low [4]. Here, we assume that *only one* neuron k, namely the one which has the largest $w(k, j)$, fires at each time step [5]. This type of "extremal dynamics" is known to organise dynamical systems into a highly adaptive (high susceptibility) critical state, [6] [7] most notably in recent models of biological evolution [8].

The dynamical process in its entirety is as follows:
An input neuron is chosen. The neuron j_m in the middle layer with the largest $w(j, i)$ is firing. Next, the output neuron k_m with the maximum $w(k, j_m)$ is firing. If the output k happens to be the desired one, **nothing** is done, otherwise $w(k_m, j_m)$ and $w(j_m, i)$ are both depressed by a fixed amount δ, which is redistributed among the other incoming synapses to the same two neurons. The redistribution can be either uniform, or to one randomly selected input.

The iterative application of this rule leads to a quick convergence to any arbitrary input-output mapping. Figure 2 shows this for a map (labelled "a") where seven inputs 1-7 are mapped to the corresponding seven outputs, 1-7, in a