3.1 Neuron Abstraction

Our discussion in Chapter 2 focussed on the microscopic behaviour of real neurons as we now understand them. Real neurons integrate hundreds or thousands of temporal signals through their dendrites. These signals modify their internal potential in a complex way that depends on the inhibitory or excitatory nature of the synapses at which the signals impinge on the cell. They transmit signals in the form of an action potential when their internal cell potential at the axon hillock exceeds a threshold of about −40 mV with respect to the external medium.

Action potentials are signals in the space-time continuum of axons.

At synapses, neurons transduce signals—electrical to chemical, and then from chemical back again to electrical. This change in the form of signals takes place with the help of complex neurotransmitters and protein based ion-specific channels. Neurotransmitters released into the extra-cellular medium selectively open or close channels in the postsynaptic membrane, thus modulating the internal cell potential of the postsynaptic neuron. Since the entire neurotransmitter released is not utilized, each synapse is associated with what we call the *synaptic efficacy*—the efficiency with which a signal is transmitted from the presynaptic to postsynaptic neuron.

3.1.1 Neuron Activations

We draw upon the neuron metaphor in order to integrate important principles of the working of real neurons into a simple mathematical neuron model. Figure 3.1 shows the jth artificial neuron that receives input signals s_i, from possibly n different sources. These signals traverse weighted pathways w_{ij}, in order to generate an internal *activation* x_j, which is a linear weighted aggregation of the impinging signals, modified by an internal threshold, θ_j:

The neuron abstraction considered by artificial neural networks represents a drastic simplification of real neurons. More recently, however, neural networks have begun to move closer to neurobiology with the introduction of pulsed neural networks which we will study in Chapter 13.

$$x_j = \sum_{i=1}^{n} w_{ij} s_i + \theta_j \tag{3.1}$$

Here, connection weights w_{ij} model the synaptic efficacies of various inter-neuron synapses. Positive weights correspond to excitatory synapses, while negative weights model inhibitory synapses. Impinging signals s_i represent mathematical abstractions of action potentials. The threshold θ_j represents the internal firing threshold of a real neuron, and the activation x_j then models the internal aggregated cell potential.

Notation: w_{ij} denotes the weight from neuron i to neuron j. A number of texts use the reverse convention.

The activation of the neuron is subsequently transformed through a *signal function* $\mathcal{S}(\cdot)$, to generate the output signal $s_j = \mathcal{S}(x_j)$ of the neuron. As we discuss in detail in Section 3.2, a signal function may typically be *binary threshold*, *linear threshold*, *sigmoidal*, *Gaussian*, or *probabilistic*.

For the sake of notational convenience we often consider the threshold θ_j as an additional weight $w_{0j} = \theta_j$ fanning into the neuron with a constant input $s_0 = +1$ (see Fig. 3.1).

This representation lends convenience in the formulation of learning algorithms which we will study in forthcoming chapters.

In biological systems, cell potentials change faster than changes in

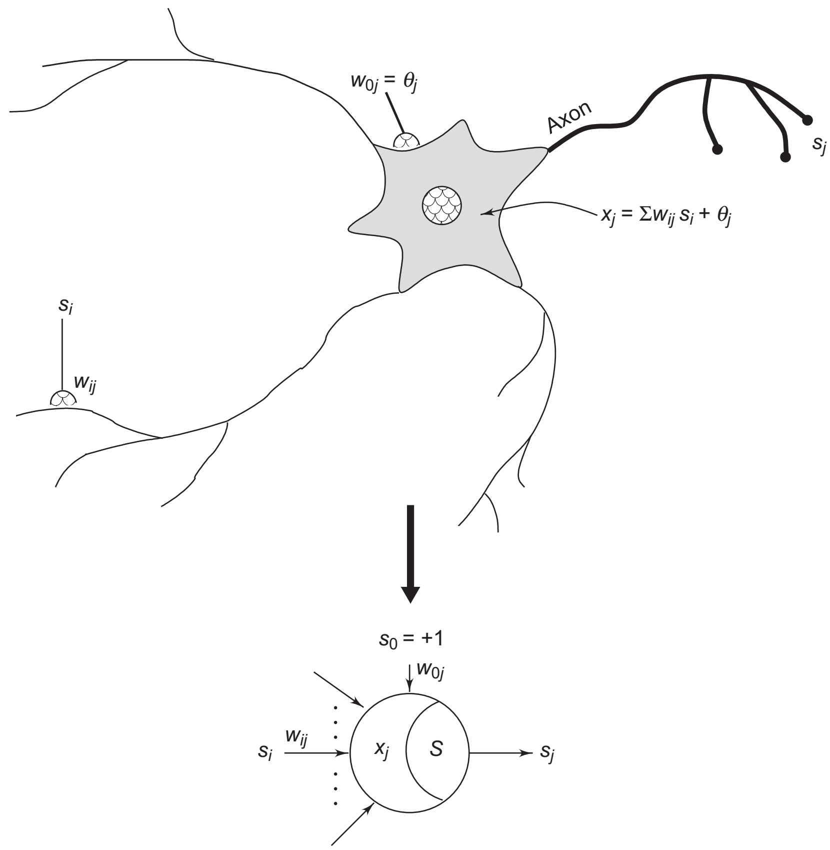


Fig. 3.1 Artificial neuron model—drawn to remind us of its biological origin

synaptic efficacies. In the brain, a thought or a reasoning process flashes across neuron ensembles in the form of fast changing patterns of internal cell potentials and action potentials. In other words, neuron activations x_j, change quickly as neurons sample and integrate impinging signals. However, the learning process—where impinging information is downloaded into memory by changing synaptic efficacies—takes place on a much longer time scale. Learning concerns itself with changes in weights w_{ij} in response to stable patterns of activity, and is a slow process. In neural networks changes in synaptic weights are implemented using a *learning algorithm*. We often refer to the rapidly changing x_j as a *short term memory* (STM) and to the slowly changing w_{ij} as *long term memory* (LTM).

3.1.2 Activations Measure Similarities

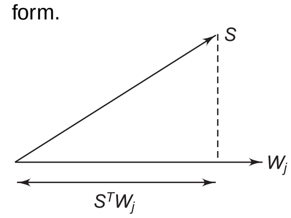
Equation (3.1) admits a linear algebraic interpretation of the process of activity aggregation at an artificial neuron. The activation x_j is simply the inner product of the impinging signal vector $S = (s_0, \ldots, s_n)^T$, with the neuronal weight vector $W_j = (w_{0j}, \ldots, w_{nj})^T$:

$$x_j = S^T W_j = \sum_{i=0}^{n} w_{ij} s_i \tag{3.2}$$

where $s_0 = +1$ and $w_{0j} = \theta_j$. This inner product represents a similarity measure, or the degree of match, between the input signal vector and the weight vector. In other words, the activation of a neuron measures the "similarity" between the input signal vector and the neuronal weight vector of the neuron in question. This formalism allows us to view the neuron as a linear *filter* in the sense that the neuron is able to discriminate between inputs that are similar to its weight vector (which tend to generate large activations), and inputs that are dissimilar to the neuron weight vector (which tend to generate small activations). Notice that if S is at right angles with W_j, the projection of S on W_j and therefore x_j is zero. Such *orthogonal* vectors will generate zero inner-products and thus zero activations. On the other hand, an input vector completely aligned with the weight vector will generate a maximal activation, and one that is aligned but in the opposite direction will generate a maximally negative activation. Always keep these points in your mind.

We have seen that a neuron filters its inputs with the help of the inner product, and measures the input-to-weight vector similarity or dissimilarity with the value of its activation. As Grossberg points out, when endowed with learning capabilities the neuron thus acquires the status of being an *adaptive filter* [194].

Note: Throughout we assume that vectors are in columnar form.



The closer S is to W_j the larger the projection or inner product or activation.