Computational Psycholinguistics

Neural Networks Review

Marshall R. Mayberry

Computerlinguistik
Universität des Saarlandes

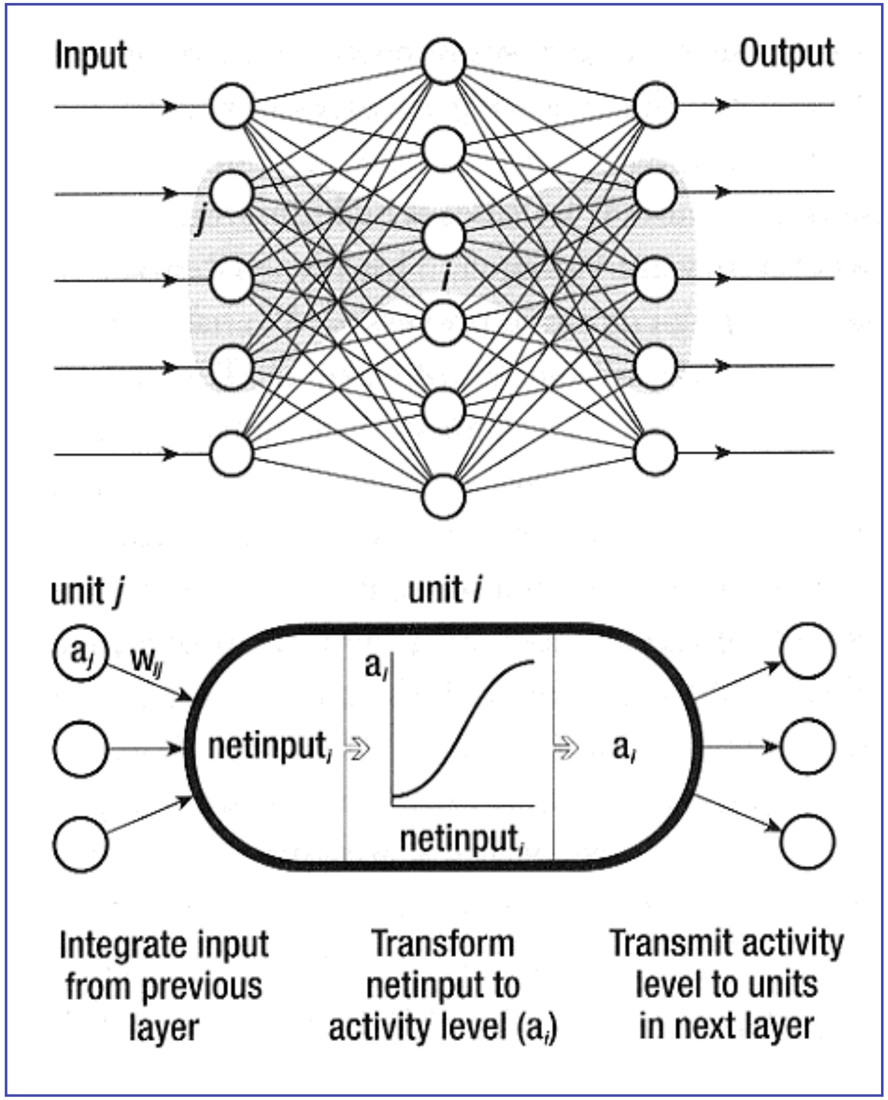
Neural network architecture

- The activation of a unit *i* is represented by the symbol a_i.
- The extent to which unit *j* influences unit i is determined by the weight w_{ij}
- The input from unit *j* to unit *i* is the product: $a_j * w_{ij}$
- For a node *i* in the network:

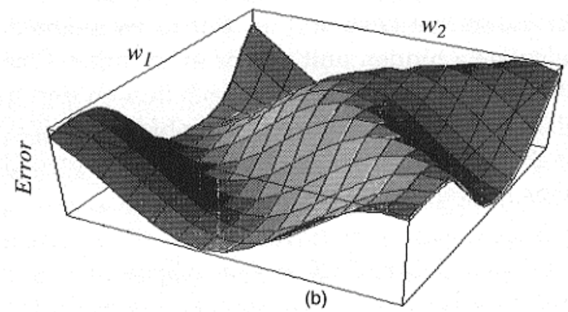
$$netinput_i = \sum_j w_{ij} a_j$$

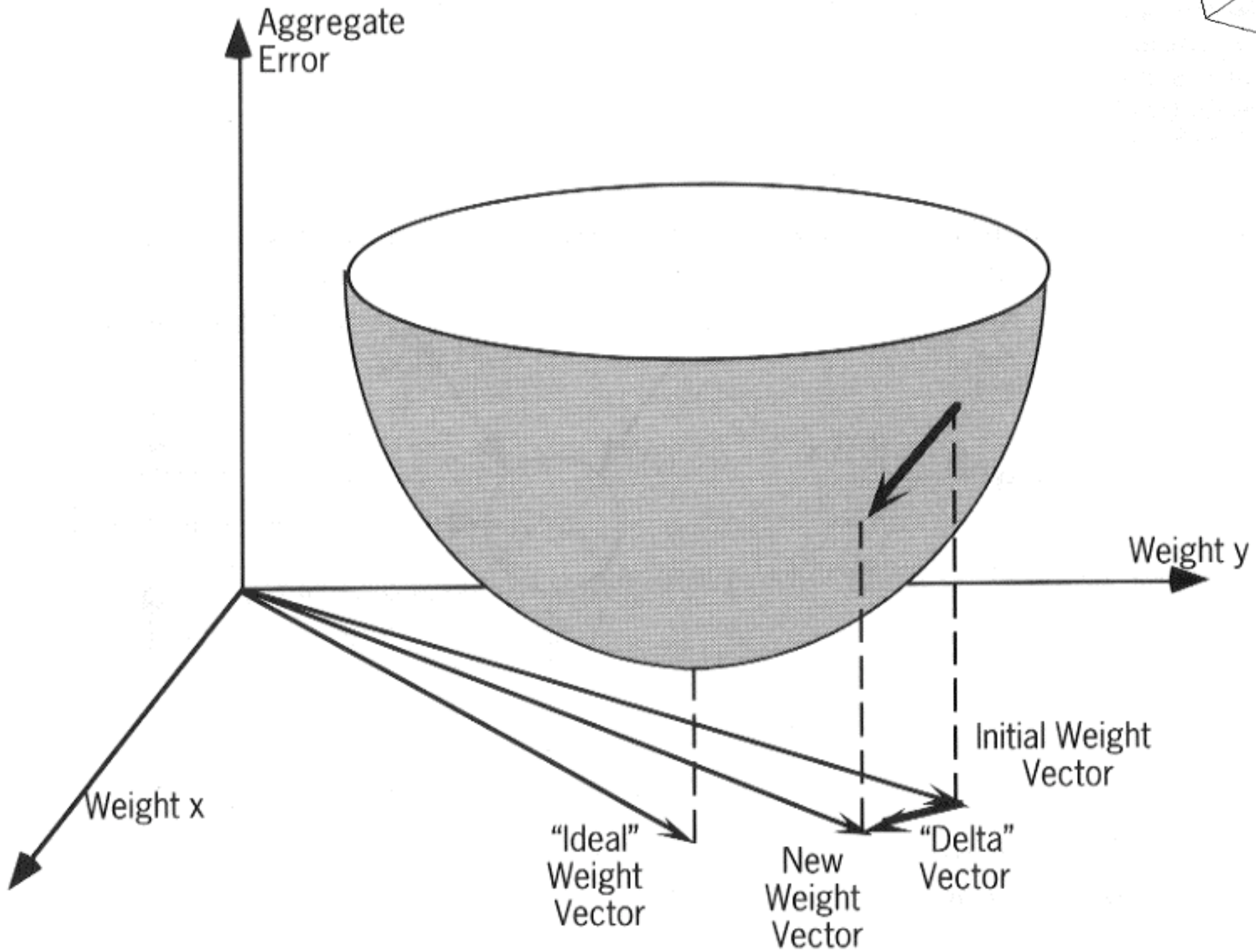
- The output activation of node *i* is determined by the activation function, e.g. the logistic:

$$a_i = \sigma(netinput_i) = \frac{1}{1 + e^{-net_i}}$$



Visualising the error “surface“

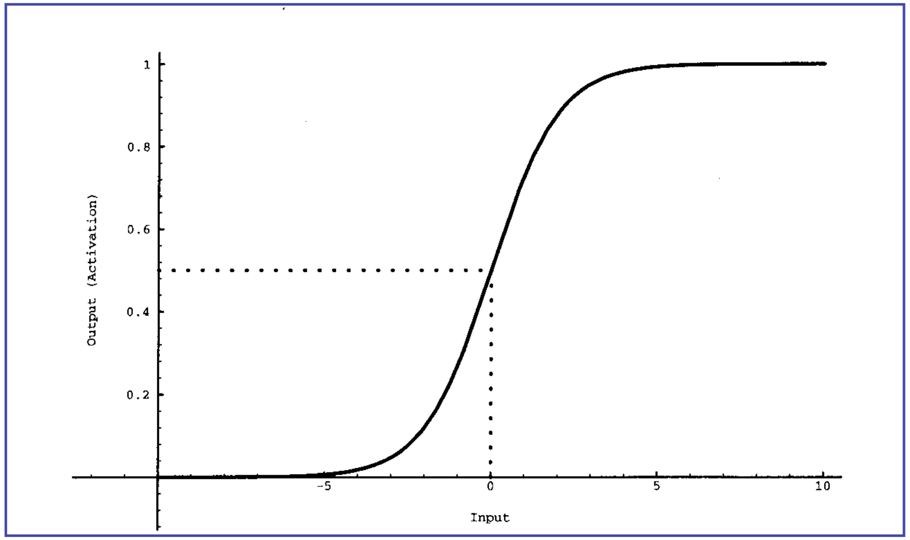




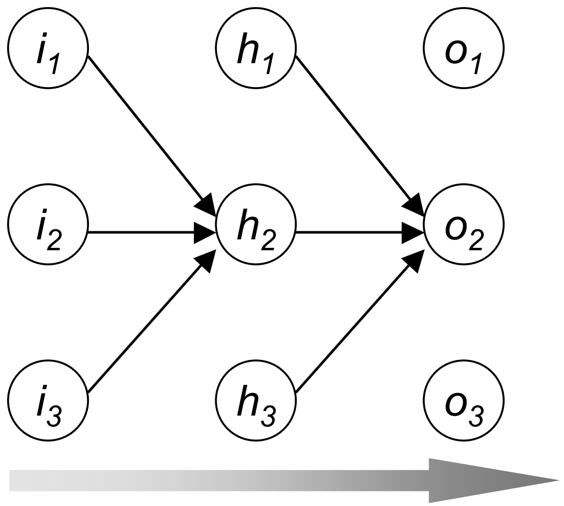
Learning with the Sigmoid activation function

- Networks with linear activation functions:
  - have mathematically well-defined learning capacities (linear algebra)
  - they are known to be limited in the kinds of problems they can solve
- The logistic, or sigmoid, function is:
  - Nonlinear: more powerful
  - More neurologically plausible
  - Less well-understood, more difficult to analyse mathematically
- Recall:

$$a_i = \sigma(net_i) = \frac{1}{1 + e^{-net_i}}$$



Backpropagation of Error



(a) Forward propagation of activity:

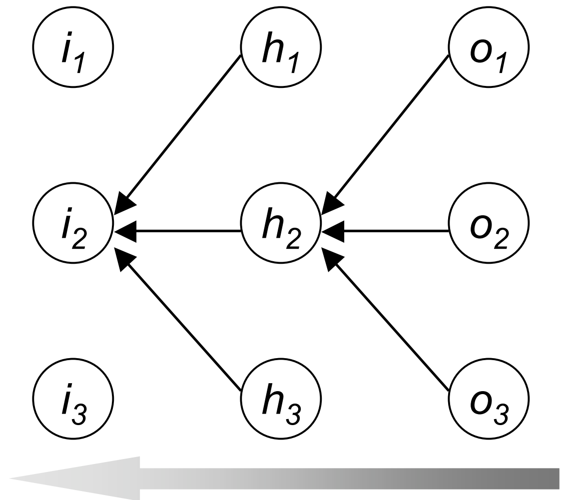
$$\text{net}_{out} = \sum w_{oh} \cdot a_{hidden}$$

$$a_{out} = f(\text{net}_{out})$$

(b) Backward propagation of error:

$$\text{err}_{hidden} = \sum w_{oh} \cdot \delta_{out}$$

$$\delta_{hidden} = f'(\text{net}_{hidden}) \cdot \text{err}_{hidden}$$



Example of Backpropagation

- Consider the following network, containing a single hidden node
- Calculate the weight changes for both layers of the network, assuming learning rate $\varepsilon = 0.3$, inputs 0.4 and 0.7, and targets 0.8 and 0.2

The generalised Delta rule:

$$\Delta w_{ij} = \varepsilon \delta_i a_j$$

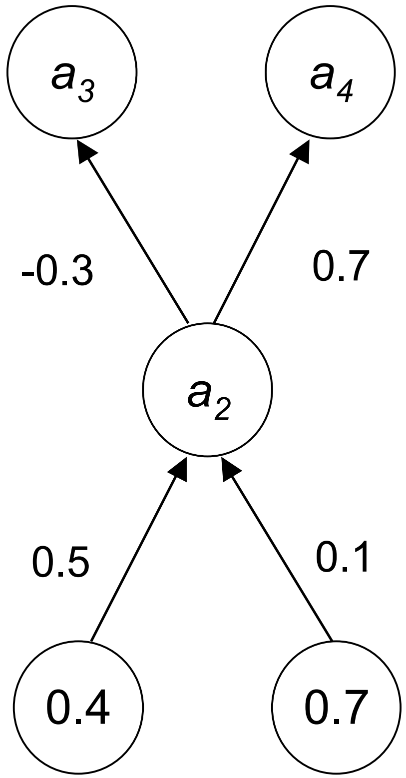
For output nodes:

$$\delta_k = \sigma'(net_k)(t_k - a_k) = a_k(1 - a_k)(t_k - a_k)$$

For hidden nodes:

$$\delta_i = \sigma'(net_i)\sum_k \delta_k w_{ki} = a_i(1 - a_i)\sum_k \delta_k w_{ki}$$

because $\sigma'(net_i) = a_i(1 - a_i)$



Forward and Backpropagation