```
        ((T . 0) (I1 . 0) (I2 . 1) (I3 . 0) (I4 . 1) (I5 . 0) (I6 . 1))
        ((T . 0) (I1 . 0) (I2 . 0) (I3 . 0) (I4 . 1) (I5 . 0) (I6 . 1))
        ((T . 0) (I1 . 0) (I2 . 1) (I3 . 1) (I4 . 0) (I5 . 1) (I6 . 1))
        ((T . 0) (I1 . 0) (I2 . 1) (I3 . 1) (I4 . 1) (I5 . 0) (I6 . 0)))))
((setq problem
      (make-learning-problem
       :attributes '((I1 0 1) (I2 0 1) (I3 0 1) (I4 0 1) (I5 0 1) (I6 0 1))
       :goals '((T 0 1))
       :examples examples)))
((setq net (perceptron-learning problem)))
((setq weights (unit-weights (first (first net)))))
((test-nn net problem)
 (= * 14))
((test-nn net problem
        '(((T . 1) (I1 . 1) (I2 . 0) (I3 . 0) (I4 . 1) (I5 . 0) (I6 . 0))
          ((T . 0) (I1 . 0) (I2 . 0) (I3 . 1) (I4 . 1) (I5 . 1) (I6 . 1))
          ((T . 1) (I1 . 1) (I2 . 1) (I3 . 0) (I4 . 0) (I5 . 1) (I6 . 0))
          ((T . 0) (I1 . 1) (I2 . 0) (I3 . 0) (I4 . 0) (I5 . 0) (I6 . 0))
          ((T . 1) (I1 . 0) (I2 . 1) (I3 . 1) (I4 . 1) (I5 . 1) (I6 . 1))))
 (= * 5)))
```

20.16 The probability p output by the perceptron is $g(\sum_j W_j a_j)$, where g is the sigmoid function. Since $g' = g(1-g)$, we have

$$\partial p/\partial W_j = g'(\sum_j W_j a_j)a_j = p(1-p)a_j$$

For a datum with actual value y, the log likelihood is

$$L = y\log p + (1-y)\log(1-p)$$

so the gradient of the log likelihood with respect to each weight is

$$\begin{aligned}\frac{\partial L}{\partial W_j} &= \frac{y}{p}\cdot\frac{\partial p}{\partial W_j} - \frac{1-y}{1-p}\cdot\frac{\partial p}{\partial W_j} \\ &= \frac{yp(1-p)a_j}{p} - \frac{(1-y)p(1-p)a_j}{1-p} = (y-p)a_j = Err \times a_j\,.\end{aligned}$$

20.17 This exercise reinforces the student's understanding of neural networks as mathematical functions that can be analyzed at a level of abstraction above their implementation as a network of computing elements. For simplicity, we will assume that the activation function is the same linear function at each node: $g(x) = cx + d$. (The argument is the same (only messier) if we allow different c_i and d_i for each node.)

a. The outputs of the hidden layer are

$$H_j = g\left(\sum_k W_{k,j}I_k\right) = c\sum_k W_{k,j}I_k + d$$

The final outputs are

$$O_i = g\left(\sum_j W_{j,i}H_j\right) = c\left(\sum_j W_{j,i}\left(c\sum_k W_{k,j}I_k + d\right)\right) + d$$