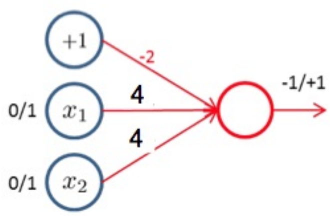
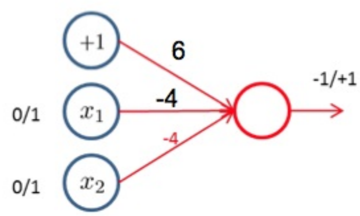
Problem Set 3
Solutions to Written Problems

1. Written problem 1
SOLUTION:

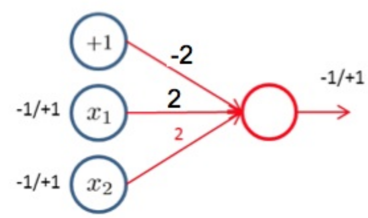
x_1 OR x_2



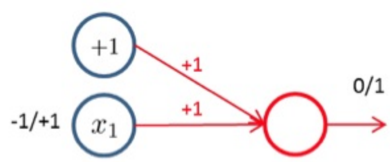
NOT(x_1 AND x_2)



x_1 AND x_2



convert to 0/1



2. Written problem 2
SOLUTION:

The likelihood function for an i.i.d. data set, $\{(x_1, t_1), ..., (x_N, t_N)\}$, under the conditional distribution:

$$p(t|x, w) = N(t|y(x, w), \beta^{-1}I)$$

is given by

$$\prod_{n=1}^{N} N(t_n|y(x_n, w), \beta^{-1}I).$$

If we take the logarithm of this, using

$$N(x|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right\},$$

we get

$$\begin{aligned}
&\sum_{n=1}^{N} \ln N(t_n|y(x_n, w), \beta^{-1}I) \\
&= -\frac{1}{2} \sum_{n=1}^{N} (t_n - y(x_n, w))^T (\beta I)(t_n - y(x_n, w)) + const \\
&= -\frac{\beta}{2} \sum_{n=1}^{N} \|t_n - y(x_n, w)\|^2 + const,
\end{aligned}$$

where 'const' comprises terms which are independent of w. The first term on the right hand side is proportional to the negative of

$$E(w) = \frac{1}{2} \sum_{n=1}^{N} \|y(x_n, w) - t_n\|^2$$

and hence maximizing the log-likelihood is equivalent to minimizing the sum-of-squares error.

3. Written problem 3
SOLUTION:

For the given interpretation of $y_k(x, w)$, the conditional distribution of the target vector for a multiclass neural network is:

$$p(t|w_1, ..., w_K) = \prod_{k=1}^{K} y_n^{t_k}$$

Thus, for a dataset of N points, the likelihood function will be:

$$p(T|w_1, ..., w_k) = \prod_{n=1}^{N} \prod_{k=1}^{K} y_{nk}^{t_{nk}}$$

Taking the negative logarithm in order to derive an error function we obtain the cross-entropy error function specified in Problem 3.

4. Written problem 4
SOLUTION:

This simply corresponds to a scaling and shifting of the binary outputs, which directly gives the activation function, using the notation from

$$y = \sigma(a) \equiv \frac{1}{1+\exp(-a)},$$

in the form

$$y = 2\sigma(a) - 1.$$

The corresponding error function can be constructed from

$$E(w) = -\sum_{n=1}^{N} \{t_n \ln y_n + (1 - t_n) \ln(1 - y_n)\}$$

by applying the inverse transform to y_n and t_n, yielding

$$E(w) = -\sum_{n=1}^{N} \frac{1+t_n}{2} \ln \frac{1+y_n}{2} + \left(1 - \frac{1+t_n}{2}\right) \ln \left(1 - \frac{1+y_n}{2}\right)$$

$$= -\frac{1}{2} \sum_{n=1}^{N} \{(1 + t_n) \ln(1 + y_n) + (1 - t_n) \ln(1 - y_n)\} + N \ln 2$$

where the last term can be dropped, since it is independent of w.

To find the corresponding activation function we simply apply the linear transformation to the logistic sigmoid given by

$$y = \sigma(a) \equiv \frac{1}{1+\exp(-a)},$$

which gives

$$y(a) = 2\sigma(a) - 1 = \frac{2}{1+e^{-a}} - 1$$

$$= \frac{1-e^{-a}}{1+e^{-a}} = \frac{e^{a/2}-e^{-a/2}}{e^{a/2}+e^{-a/2}}$$

$$= \tanh(a/2).$$