4.8)

To change the output function from the sigmoid to tanh, we need only make a couple changes.

The first is to update (T4.3) we change from:

To

We must also update (T4.4) in a similar fashion to:

4.10)

So taking the error with respect to a single document rather than the entire network:

The gradient descent update rule is:

And now we substitute in the error equation and perform the derivative:

Since the first term’s sum is over only the outputs, we can simply set k=q. The second term disappears because there is no o term in it:

And now we need to calculate . Note that o is the sigmoid function and the second term that disappeared in the previous derivation does not disappear here. This term is actually the same as the net for the output units:

Now we combine these and obtain:

So

And the gradient descent rule is: