



Figure 6.12 A variety of nonlinear activation functions.

biological neurons. A closely related function is \tanh , which is defined by

$$\tanh(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}, \quad (6.14)$$

which is plotted in [Figure 6.12\(a\)](#). This function differs from the logistic sigmoid by a linear transformation of its input and its output values, and so for any network with logistic-sigmoid hidden-unit activation functions, there is an equivalent network with \tanh activation functions. However, when training a network, these are not necessarily equivalent because for gradient-based optimization, the network weights and biases need to be initialized, and so if the activation functions are changed, then the initialization scheme must be adjusted accordingly. A ‘hard’ version of the \tanh function (Collobert, 2004) is given by

$$h(a) = \max(-1, \min(1, a)) \quad (6.15)$$

and is plotted in [Figure 6.12\(b\)](#).

A major drawback of both the logistic sigmoid and the \tanh activation functions is that the gradients go to zero exponentially when the inputs have either large positive or large negative values. We will discuss this ‘vanishing gradients’ issue later,

Exercise 6.4

Section 7.4.2