Contents

Chapter 1

1.1

1.1.1 Estimating class probabilities in multiclass classification via the softmax function

The softmax function is a soft form of the argmax function; instead of giving a single class index, it provides the probability of each class. Therefore, it allows us to compute meaningful class probabilities in multiclass settings (multinomial logistic regression).

In softmax, the probability of a particular sample with net input z belonging to the ith class can be computed with a normalization term in the denominator, that is, the sum of the exponentially weighted linear functions:

$$p(z) = \sigma(z) = \frac{e^{z_i}}{\sum_{j=1}^{M} e^{z_j}} \tag{1.1}$$

1.1.2 Broadening the output spectrum using a hyperbolic tangent

Another sigmoidal function that is often used in the hidden layers of artificial NNs is the hyperbolic tangent (commonly known as tanh), which can be interpreted as a rescaled version of the logistic function:

$$\begin{aligned} \sigma_{logistic}(z) &= \frac{1}{1+e^{-z}} \\ \sigma_{tanh}(z) &= 2 \times \sigma_{logistic}(2z) - 1 = \frac{e^{z}-e^{-z}}{e^{z}+e^{-z}} \end{aligned} \tag{1.2}$$

The advantage of the hyperbolic tangent over the logistic function is that it has a broader output spectrum ranging in the open interval $(-1, 1)$, which can improve the convergence of the backpropagation algorithm.

Note that using torch.sigmoid(x) produces results that are equivalent to torch. nn.Sigmoid()(x). torch.nn.Sigmoid is a class to which you can pass in parameters to construct an object in order to control the behavior. In contrast, torch.sigmoid is a function.

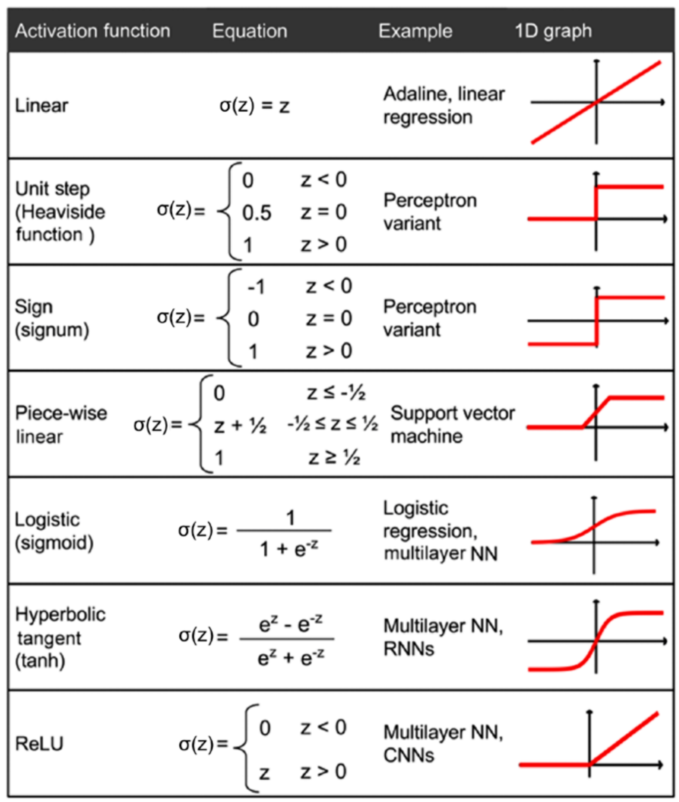

Figure 1.1: The activation functions covered