

Figure 29-4. One-hot encoding

In the final layer of the neural network, just before the output layer, an activation function called the softmax (same as discussed under “Logistic Regression”) is applied to transform the activations to the probability that the example belongs to one of the output classes.

The purpose of applying one-hot encoding to the labels of the dataset is to represent the output as a vector of distinct classes with the probability that an example in the training dataset belongs to any one of the output categories.

The Backpropagation Algorithm

Backpropagation is the process by which we train the neural network to improve its prediction accuracy. To train the neural network, we need to find a mechanism for adjusting the weights of the network; this in turn affects the value of the activations within each neuron and consequently updates the value of the predicted output layer. The first time we run the feedforward algorithm, the activations at the output layer are most likely incorrect with a high error estimate or cost function.

The goal of backpropagation is to repeatedly go back and adjust the weights of each preceding neural layer and perform the feedforward algorithm again until we minimize the error made by the network at the output layer (see Figure 29-5).

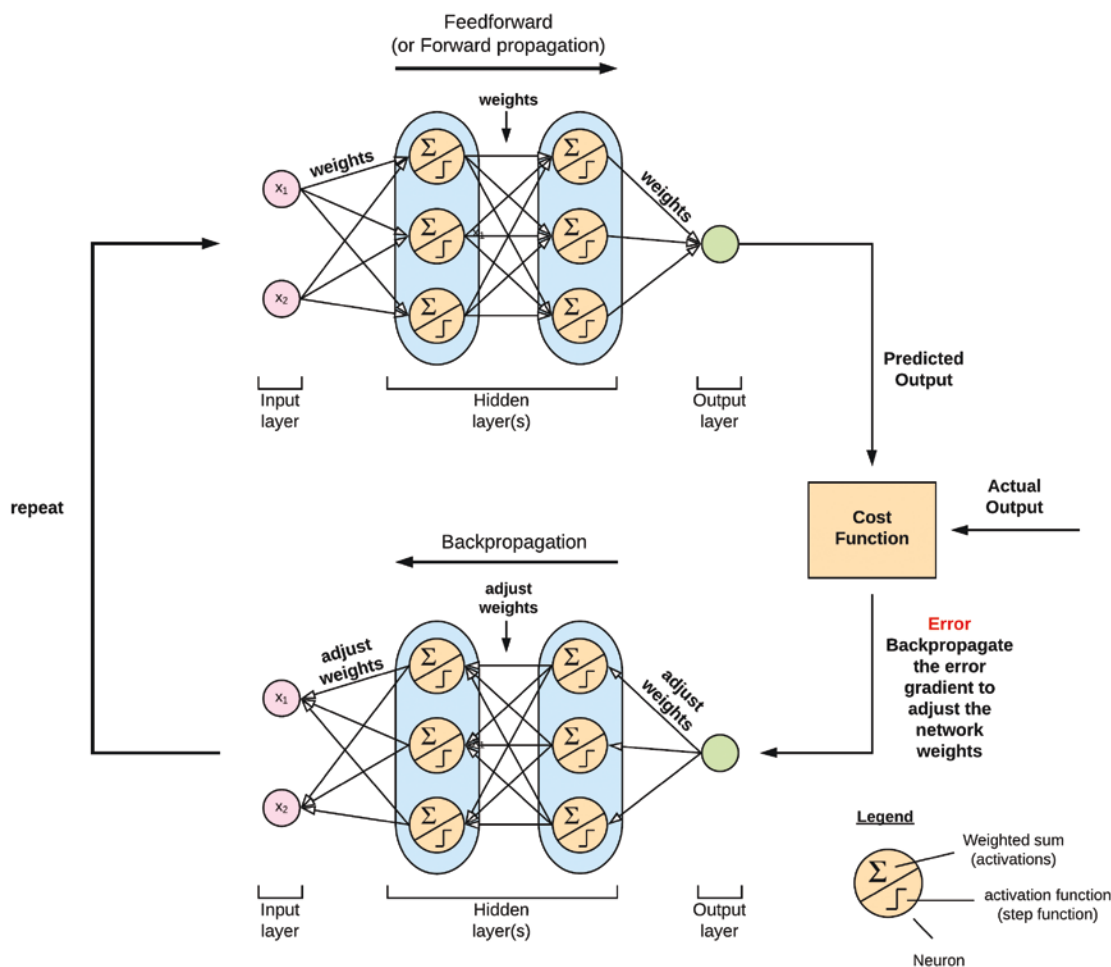


Figure 29-5. *Backpropagation*

The backpropagation algorithm works by computing the cost function at the output layer by comparing the predicted output of the neural network with the actual outputs from the dataset. It then employs gradient descent (earlier discussed in Chapter 16) to calculate the gradient of the cost function using the weights of the neurons at each successive layer and update the weights propagating back through the network.

Activation Functions

Up till now, we have mentioned activation functions. Now let’s go a bit deeper into what activation functions are and why do we have them.