The Multilayer Perceptron (MLP)

The multilayer perceptron (MLP) is the fundamental example of a deep neural network. The architecture of a MLP consists of multiple hidden layers to capture more complex relationships that exist in the training dataset. Another name for the MLP is the deep feedforward neural network (DFN). An illustration of an MLP is shown in Figure 31-1.
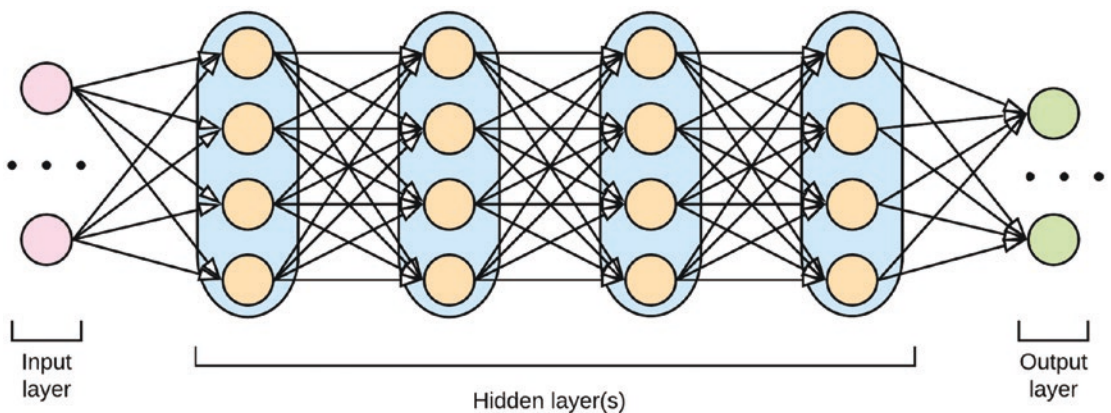

Figure 31-1. *Deep feedforward neural network*

The Concept of Hierarchies

The more the number of hidden layers in a neural network, the deeper the network becomes. Deep networks are able to learn more sophisticated representations of the inputs. The concept of hierarchical representation is when each layer learns a set of features that describe the input and hierarchically pass that information across the hidden layers. Initially, the hidden layers closer to the input layer learn a simple set

of features, which then grow to increasingly complex features as information flows to deeper layers of the network, to capture the mapping between the inputs and the target. See Figure 31-2.
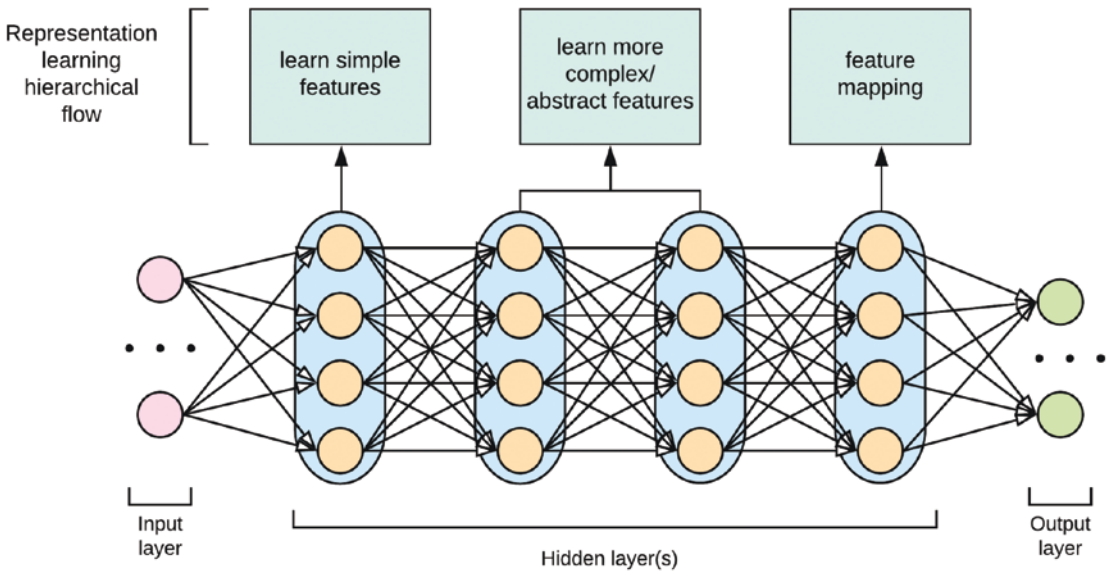

Figure 31-2. *Hierarchical learning*

Choosing the Number of Hidden Layers: Bias/Variance Trade-Off

From experience, increasing the number of hidden layers may improve the representational quality of the network; however, arbitrarily increasing the number of hidden layers in your network design can have detrimental effects on the overall network performance with respect to generalizing to unseen observations. This is because the neural network will learn more closely the irreducible errors inherent in the training dataset and will fail to generalize to new examples.

Appropriate caution should be taken when selecting the number of hidden layers to avoid overfitting. Regularization techniques for neural networks such as Tikhonov regularization, Dropout, or early stopping are different methods of mitigating overfitting. Regularization for neural networks will be covered in more detail in a later section.

Empirically, one hidden layer will produce good results for simple learning problems, but if the number of output classes increases or there exists a high degree