From the unrolled graph of the recurrent neural network, we can observe how the input into the recurrent layer includes the output of the previous time step $t-1$ in addition to the current input at time step t. This architecture of the recurrent neuron is central to how the recurrent neural network learns from past events or past sequences.

Up until now, we have seen that the recurrent neuron captures information from the past by storing memory or state in its memory cell. The recurrent neuron can have a much more complicated memory cell (such as the GRU or LSTM cell) than the basic RNN cell as illustrated in the images so far, where the output at time instant $t-1$ holds the memory.

Basic Recurrent Neural Network

Earlier on, we mentioned that when a recurrent network is unfolded, we can see how information flows from one recurrent layer to the other. Further, we noted that the sequence length of the dataset determines the number of recurrent layers. Let's briefly illustrate this point in Figure 36-4. Suppose we have a time series dataset of ten layers, for each row sequence in the dataset, we will have ten layers in the recurrent network system.

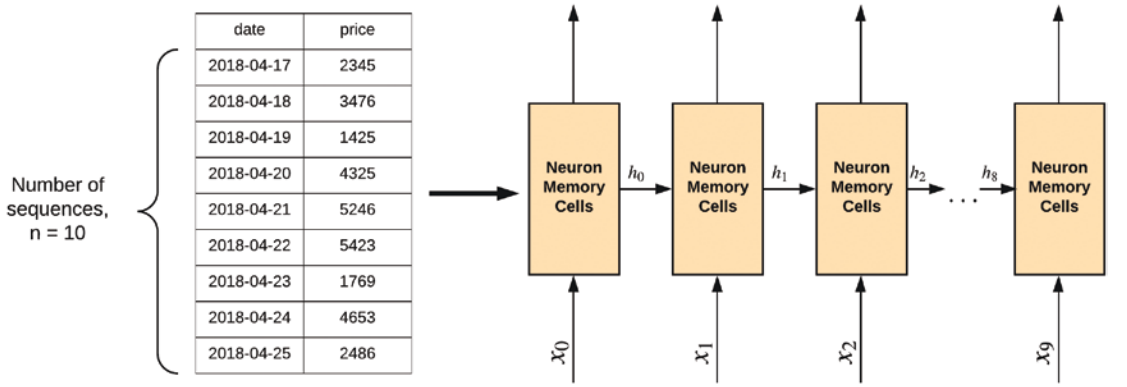

Figure 36-4. *Dataset to layers*

At this point, we must firmly draw attention to the fact that the recurrent layer does not comprise of just one neuron cell, but it is instead a set of neurons or neuron cells as shown in Figure 36-5. The choice of the number of neurons in a recurrent layer is a design decision when composing the network architecture.

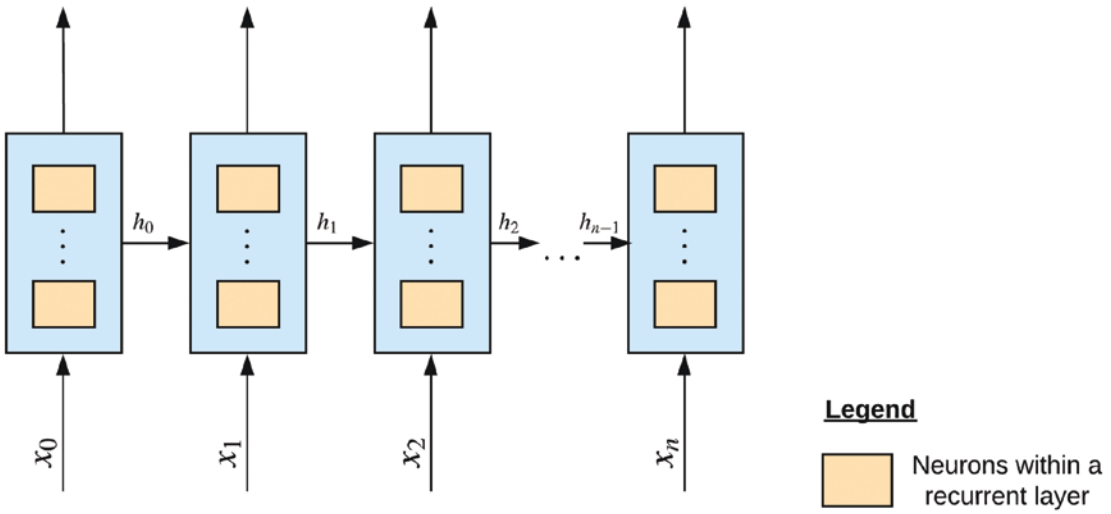

Figure 36-5. *Neurons in a recurrent layer*

Each neuron in a recurrent layer receives as input the output of the previous layer and its current input. Hence, the neurons each have two weight vectors. Again, just like other neurons, they perform an affine transformation of the inputs and pass it through a non-linear activation function (usually the hyperbolic tangent, tanh). Still, within the recurrent layer, the output of the neurons is moved to a dense or fully connected layer with a softmax activation function for outputting the class probabilities. This operation is illustrated in Figure 36-6.

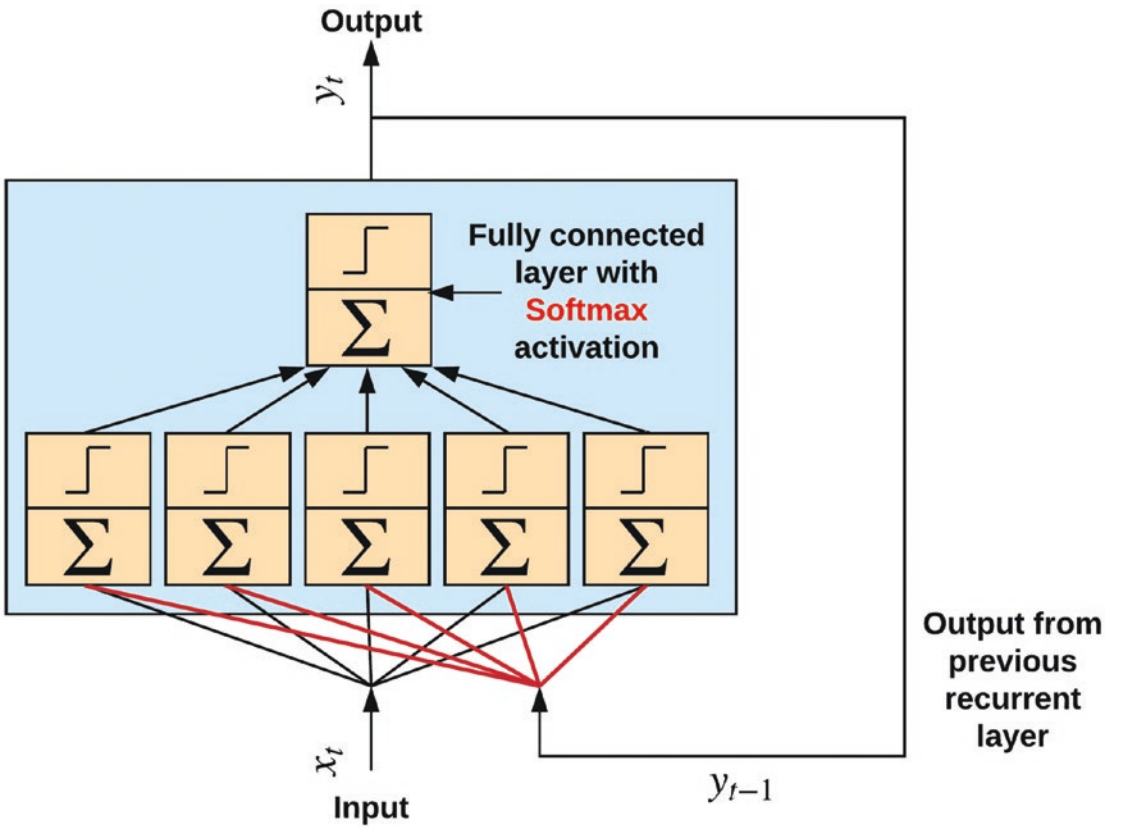

Figure 36-6. *Computations within a recurrent layer*

Recurrent Connection Schemes

There are two main schemes for forming recurrent connections from one recurrent layer to another. The first is to have recurrent connections between hidden units, and the other is recurrent connections between the hidden unit and the output of the previous layer. The different schemes are visually illustrated in Figure 36-7.