Convolutional Layer

A convolutional network assumes special spatial structure in its input. In particular, it assumes that inputs that are close to each other spatially are semantically related. This assumption makes most sense for images, since pixels close to one another are likely semantically linked. As a result, convolutional layers have found wide use in deep architectures for image processing. This concept is illustrated in Figure 1-2.

Just like fully connected layers transform lists to lists, convolutional layers transform images into images. As a result, convolutional layers can be used to perform complex image transformations, such as applying artistic filters to images in photo apps.

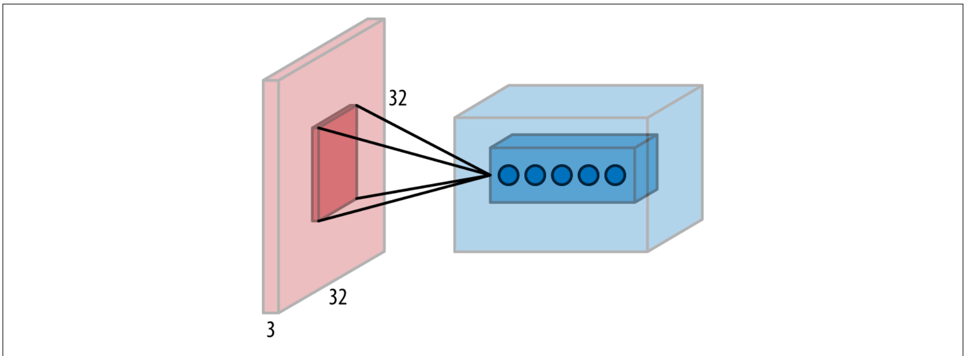

Figure 1-2. A convolutional layer. The red shape on the left represents the input data, while the blue shape on the right represents the output. In this particular case, the input is of shape (32, 32, 3). That is, the input is a 32-pixel-by-32-pixel image with three RGB color channels. The highlighted region in the red input is a "local receptive field," a group of inputs that are processed together to create the highlighted region in the blue output.

Recurrent Neural Network Layers

Recurrent neural network (RNN) layers are primitives that allow neural networks to learn from sequences of inputs. This layer assumes that the input evolves from step to step following a defined update rule that can be learned from data. This update rule presents a prediction of the next state in the sequence given all the states that have come previously. An RNN is illustrated in Figure 1-3.

An RNN layer can learn this update rule from data. As a result, RNNs are very useful for tasks such as language modeling, where engineers seek to build systems that can predict the next word users will type from history.

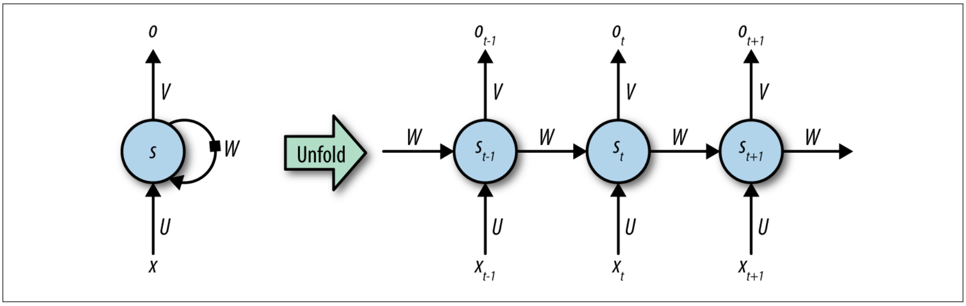

Figure 1-3. A recurrent neural network (RNN). Inputs are fed into the network at the bottom, and outputs extracted at the top. W represents the learned transformation (shared at all timesteps). The network is represented conceptually on the left and is unrolled on the right to demonstrate how inputs from different timesteps are processed.

Long Short-Term Memory Cells

The RNN layers presented in the previous section are capable of learning arbitrary sequence-update rules in theory. In practice, however, such layers are incapable of learning influences from the distant past. Such distant influences are crucial for performing solid language modeling since the meaning of a complex sentence can depend on the relationship between far-away words. The long short-term memory (LSTM) cell is a modification to the RNN layer that allows for signals from deeper in the past to make their way to the present. An LSTM cell is illustrated in Figure 1-4.

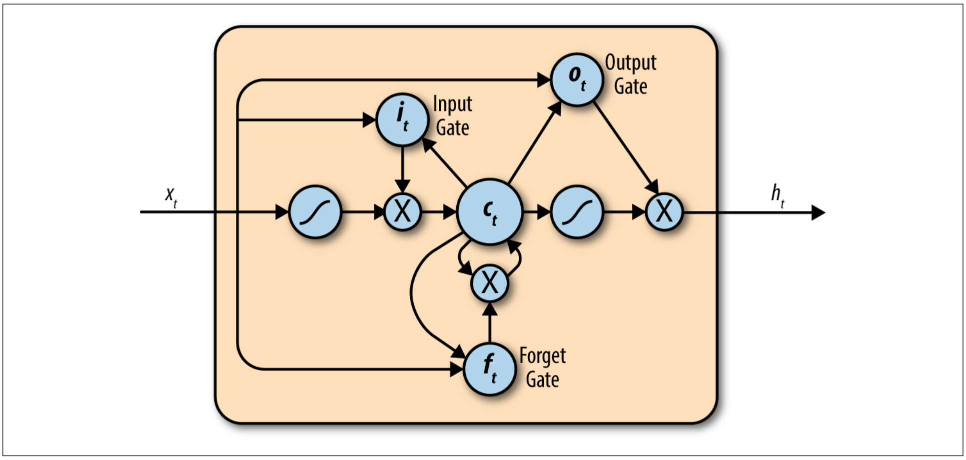

Figure 1-4. A long short-term memory (LSTM) cell. Internally, the LSTM cell has a set of specially designed operations that attain much of the learning power of the vanilla RNN while preserving influences from the past. Note that the illustration depicts one LSTM variant of many.