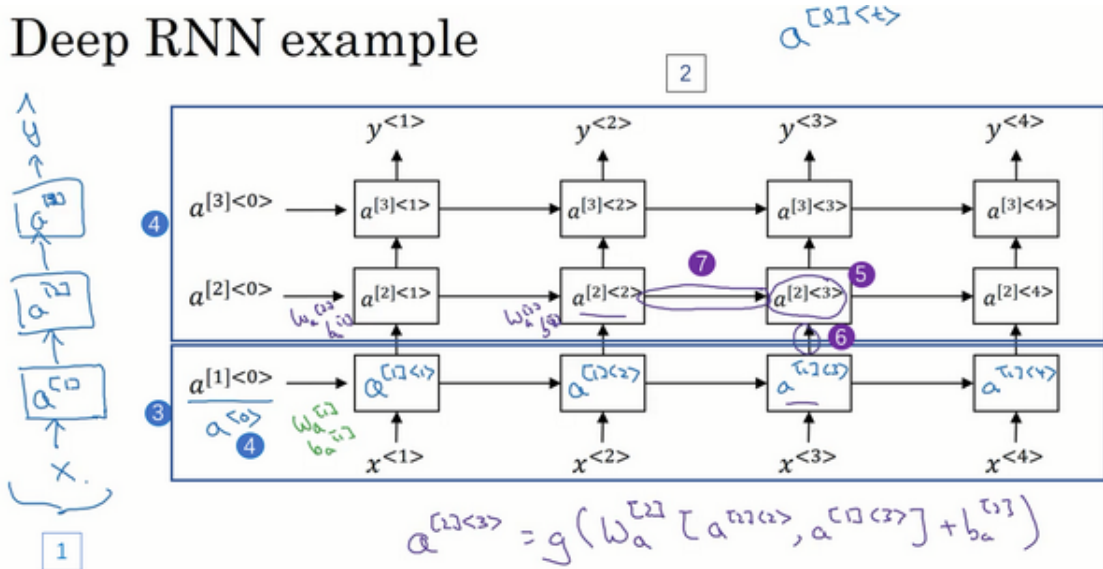


1.12 深层循环神经网络（Deep RNNs）

目前你学到的不同 RNN 的版本，每一个都可以独当一面。但是要学习非常复杂的函数，通常我们会把 RNN 的多个层堆叠在一起构建更深的模型。这节视频里我们会学到如何构建这些更深的 RNN。

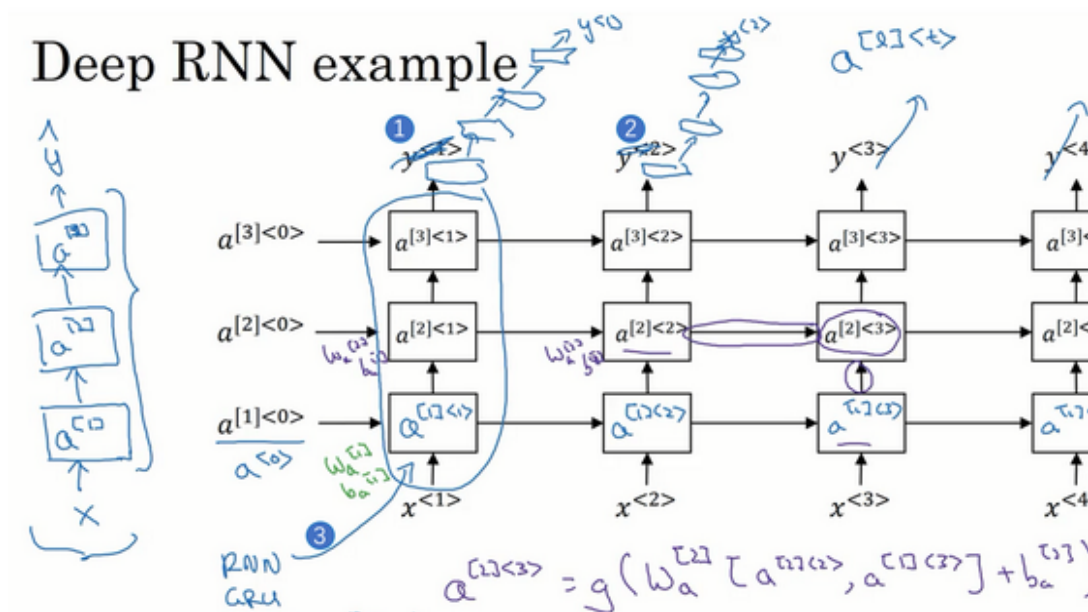
一个标准的神经网络，首先是输入 x ，然后堆叠上隐含层，所以这里应该有激活值，比如说第一层是 $a^{[1]}$ ，接着堆叠上下一层，激活值 $a^{[2]}$ ，可以再加一层 $a^{[3]}$ ，然后得到预测值 \hat{y} 。深层的 RNN 网络跟这个有点像，用手画的这个网络（下图编号 1 所示），然后把它按时间展开就是了，我们看看。



这是我们一直见到的标准的 RNN（上图编号 3 所示方框内的 RNN），只是我把这里的符号稍微改了一下，不再用原来的 $a^{<0>}$ 表示 0 时刻的激活值了，而是用 $a^{[1]<0>}$ 来表示第一层（上图编号 4 所示），所以现在我们用 $a^{[l]<t>}$ 来表示第 l 层的激活值，这个 $<t>$ 表示第 t 个时间点，这样就可以表示。第一层第一个时间点的激活值 $a^{[1]<1>}$ ，这 $(a^{[1]<2>})$ 就是第一层第二个时间点的激活值， $a^{[1]<3>}$ 和 $a^{[1]<4>}$ 。然后我们把这些（上图编号 4 方框内所示的部分）堆叠在上面，这就是一个有三个隐层的新的网络。

我们看个具体的例子，看看这个值 $(a^{[2]<3>})$ ，上图编号 5 所示）是怎么算的。激活值 $a^{[2]<3>}$ 有两个输入，一个是从下面过来的输入（上图编号 6 所示），还有一个是从左边过来的输入（上图编号 7 所示）， $a^{[2]<3>} = g(W_a^{[2]}[a^{[2]<2>}, a^{[1]<3>}] + b_a^{[2]})$ ，这就是这个激活值的计算方法。参数 $W_a^{[2]}$ 和 $b_a^{[2]}$ 在这一层的计算里都一样，相对应地第一层也有自己的参数 $W_a^{[1]}$

和 $b_a^{[1]}$ 。



对于像左边这样标准的神经网络，你可能见过很深的网络，甚至于 100 层深，而对于 **RNN** 来说，有三层就已经不少了。由于时间的维度，**RNN** 网络会变得相当大，即使只有很少的几层，很少会看到这种网络堆叠到 100 层。但有一种会容易见到，就是在每一个上面堆叠循环层，把这里的输出去掉（上图编号 1 所示），然后换成一些深的层，这些层并不水平连接，只是一个深层的网络，然后用来预测 $y^{<1>}$ 。同样这里（上图编号 2 所示）也加上一个深层网络，然后预测 $y^{<2>}$ 。这种类型的网络结构用的会稍微多一点，这种结构有三个循环单元，在时间上连接，接着一个网络在后面接一个网络，当然 $y^{<3>}$ 和 $y^{<4>}$ 也一样，这是一个深层网络，但没有水平方向上的连接，所以这种类型的结构我们会见得多一点。通常这些单元（上图编号 3 所示）没必要非是标准的 **RNN**，最简单的 **RNN** 模型，也可以是 **GRU** 单元或者 **LSTM** 单元，并且，你也可以构建深层的双向 **RNN** 网络。由于深层的 **RNN** 训练需要很多计算资源，需要很长的时间，尽管看起来没有多少循环层，这个也就是在时间上连接了三个深层的循环层，你看不到多少深层的循环层，不像卷积神经网络一样有大量的隐含层。

这就是深层 **RNN** 的内容，从基本的 **RNN** 网络，基本的循环单元到 **GRU**，**LSTM**，再到双向 **RNN**，还有深层版的模型。这节课后，你已经可以构建很不错的学习序列的模型了。