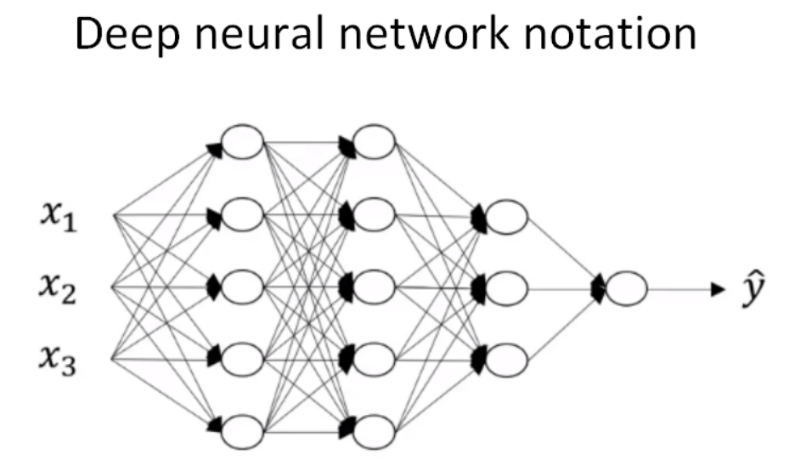
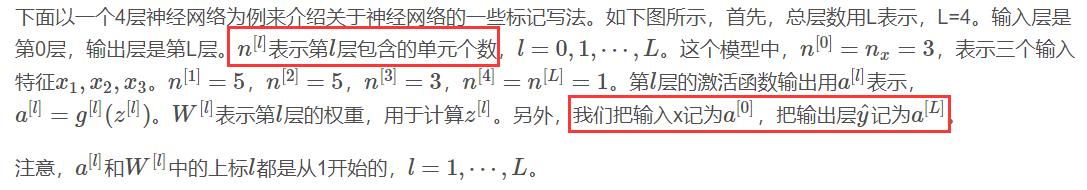
## 深层神经网络

**1.深L层神经网络**



输入层是0层，输出层是L层



**2.深层网络中的正向传播**

**3.正确的矩阵维数**

**4.为什么深度这么有理**

神经网络第一层所做的事就是从原始图片中提取出人脸的轮廓与边缘，即边缘检测。这样每个神经元得到的是一些边缘信息。神经网络第二层所做的事情就是将前一层的边缘进行组合，组合成人脸一些局部特征，比如眼睛、鼻子、嘴巴等。再往后面，就将这些局部特征组合起来，融合成人脸的模样。可以看出，随着层数由浅到深，神经网络提取的特征也是从边缘到局部特征到整体，由简单到复杂。可见，如果隐藏层足够多，那么能够提取的特征就越丰富、越复杂，模型的准确率就会越高

语音识别模型也是这个道理。浅层的神经元能够检测一些简单的音调，然后较深的神经元能够检测出基本的音素，更深的神经元就能够检测出单词信息。如果网络够深，还能对短语、句子进行检测。记住一点，神经网络从左到右，神经元提取的特征从简单到复杂。特征复杂度与神经网络层数成正相关。特征越来越复杂，功能也越来越强大。

除了从提取特征复杂度的角度来说明深层网络的优势之外，深层网络还有另外一个优点，就是能够减少神经元个数，从而减少计算量。

**5.参数VS超参数**

而超参数则是例如学习速率α，训练迭代次数N，神经网络层数L，各层神经元个数 n[l]，激活函数等。之所以叫做超参数的原因是它们决定了参W[l]和b[l]的值。

如何设置最优的超参数是一个比较困难的、需要经验知识的问题。通常的做法是选择超参数一定范围内的值，分别代入神经网络进行训练，测试cost function随着迭代次数增加的变化，根据结果选择cost function最小时对应的超参数值。这类似于validation的方法。

**6.深度学习和大脑的关联性**

关联不大