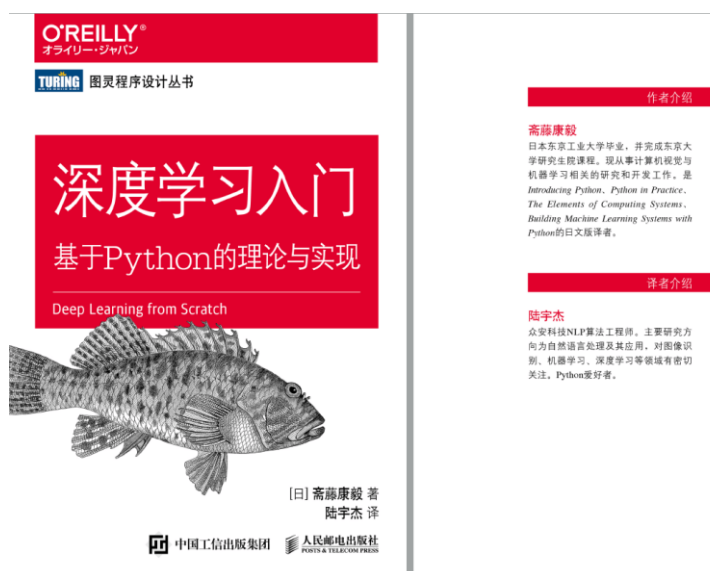


# 基于python的深度学习入门

## Deep Learning from Scratch

这是本人在关于深度学习领域所阅读的第一本书，即日本程序员斋藤康毅著，陆宇杰译作的《深度学习入门：基于python的理论与实现》。全书用非常通俗易懂的语言讲解了人工神经网络是什么、怎样运行、如何搭建。根据作者的初衷，读者在读完这本书之后应当能够利用python自行完成神经网络的构筑。



下面是我对核心章节的“极其浓缩”的概括：

## 第二章——感知机。

感知机是具有输入和输出的算法，通过设定权重、偏置等参数，来约束从输入到输出的中间过程。通过感知机可以表示逻辑电路，但是单层感知机只能表示线性空间（对应“与或非”逻辑电路），多层感知机才可以表示非线性空间（以“异或门”电路为代表）。从理论上来说，通过组合感知机就可以表示计算机。

### 第三章——神经网络。

在感知机（朴素感知机，单层神经网络）中，权重的设定是由人工完成的；而神经网络能够自动地从数据中学习到合适的权重参数。相对于某一单层来说，激活函数将输入信号通过某种规则转化为输出信号，它正是连接感知机和神经网络的桥梁。在激活函数上，感知机选择了阶跃函数，而神经网络（本书主要）选择了 Sigmoid 函数和 ReLU 函数（平滑性）。值得一提的是，多层神经网络的优势必须要通过非线性的激活函数才能体现出来。

对于整体的输出层来说，激活函数的选择更要依据问题的性质进行选择。通常来说，回归问题用恒等函数，分类问题用 Softmax 函数。其中，Softmax 具有各元素之和为 1 的性质，可以表示“概率”。为了防止计算时分母的溢出，我们常常对所有元素同加或同减某一个数再进行计算。

在多层神经网络的实现中，我们常常可采用批处理来缩短处理时间。因为批处理综合考虑了数据传输和计算，当然也这归功于 NumPy 多维数组的巧妙。

### 第四章——神经网络的学习。

与创造算法解决问题不同，神经网络从数据中进行“学习”，对问题的性质和种类没有依赖性。神经网络的学习是获取最优权重参数的过程。但由于“识别精度”的数据往往较为离散，对微小的参数变

化不敏感，因此我们以“损失函数”为基准，将“学习”变为不断寻找权重参数、使损失函数尽可能小的过程。

损失函数是神经网络性能“恶劣程度”的指标，一般用均方误差或者交叉熵误差（主）来表示。对于如何改变参数来减小损失函数的问题，本章中采用数值微分的形式代替求导，让权重参数在梯度方向作微小改变来更新。

“学习”的过程中利用的数据，可以分为“训练数据”和“测试数据”（也称“监督数据”）。“学习”的过程可分三步走：第一步随机选用训练数据组成 mini-batch，推进计算损失函数；第二步计算损失函数的梯度，对各个参数进行更新；第三步，即重复步骤一、二。另外，每完成一个 epoch，可以结合测试数据对权重参数做评价，来保证其泛化能力，防止过拟合。

## 第五章——误差反向传播法

本章引入了两个概念——误差反向传播法、用“层”将神经网络模块化。

计算图将一个复杂的计算分割成简单的局部计算，保存了计算过程中的所有“中间量”。更重要的是，蕴含“链式法则”的反向传播能高效地计算导数，可以取代之前利用数值微分求导的方法。但这并不意味着数值微分的方法就不再被使用了，实际上，在确认误差反向传播法的实现是否正确时，仍需要用到数值微分的方法（该操作称为“梯度确认”）。

以“买苹果”的例子作引子，书中实现了加法层和乘法层，后来逐步推进又实现了其他层，如激活函数上实现了 ReLU 层和 Sigmoid 层，矩阵计算上实现了 Affine 层，输出上完成了 Softmax-with-Loss 层。像这样使用层来进行模块化，神经网络中可以自由地组装层，轻松构建出想要的网络。

## 第六章——与学习相关的技巧

本章意图引入神经网络的学习中的一些重要观点，使用本章介绍的这些方法，可以高效地进行神经网络的学习，提高识别精度。

在最优权重参数的最优化方法中，由于随机梯度下降法 (SGD) 的梯度方向实际并不指向最小值的方向，所以本书引入 Momentum、AdaGrad、Adam 三种方法来取代 SGD，在速度和精度上都有所提升。

在权重参数的初始值的选取上，一方面为了防止“权重均一化”，不可以将初始值设为 0，必须随机生成；另一方面初始值过大或过小分别会导致“梯度消失”和“表现力受限”的问题。因此本书建议根据不同的激活函数选取不同的权重初始值的推荐值（ReLU 函数对应 He 初始值，Sigmoid 函数等 S 型曲线对应 Xavier 初始值）。

Batch-Normalization 强制调整了各层的激活值分布，使其拥有适当的广度。将这样的处理插入到激活函数的前面（或者后面），可以减小数据分布的偏向。总的来说，该算法具有使学习快速进行、降低对初始值的依赖性、抑制过拟合等优点。

正则化是一个抑制过拟合的过程，本书介绍了权值衰减 (weight

decay) 和 Dropout 方法。逐渐缩小“好值”存在的范围是搜索超参数的一个有效方法。

上面是本人认为的全书最精华的部分。另外，本书还包括了卷积神经网络的运行原理和未来技术发展的前瞻等内容，亦值得通读。