

145 | 基于深度学习的计算机视觉技术（三）：深度学习模型的优化

2018-09-05 洪亮劼

AI技术内参

[进入课程 >](#)



讲述：初明明

时长 07:25 大小 3.40M



在上一讲的分享里，我们聊了应用到图像上的一些最基本的深度学习模型，主要讨论了前馈神经网络和卷积神经网络的定义，以及它们在图像处理上的应用。

今天，我们从**优化**的角度来讨论，如何对深度学习模型进行训练。可以说，模型优化是成功利用深度学习模型的关键步骤。

一般机器学习模型的优化

要想了解深度学习模型的优化，我们首先得来看一看一般机器学习模型的优化。先了解一些基本的步骤，我们在讨论深度学习模型优化的时候就能更容易地看清事物的本质。

在开始说模型优化之前，要说一点需要注意的问题，我觉得这一点对于初学者来说尤为重要，那就是**要区分模型、目标函数和优化过程**。这三个实体相互关联而且相互影响，我们需要对每一个实体都有一个清晰的理解。

我们以线性模型作为例子，来感受下这三个实体的关系。

我们说一个模型是线性**模型**，是指我们期望利用一组特征（Feature）来对一个输出反馈进行线性预测。这里的线性其实主要是指参数针对于反馈变量而言是线性的。

需要注意的是，线性模型是一个数学模型。线性模型的设置本身并没有限定这个模型的参数（也就是模型的系数）是如何得来的，也就是数学模型本身的设置和得到参数的过程往往是互相独立的。我们把得到参数的过程叫作**模型训练**或者简称为**模型优化过程**。

对于线性模型而言，我们常常利用最小二乘法来构造参数学习的**目标函数**。在最小二乘法的目标函数下，一般情况下，我们既可以得到一个“解析解”（Closed Form Solution），也能通过例如梯度下降的方法来进行数值优化。

对模型、目标函数和优化过程这几个概念有了清晰的认识后，那具体的模型优化过程是怎样的呢？

这里，我们就总结一下一般机器学习模型，主要是简单模型的优化过程。

模型优化的第一步就是选择目标函数。总的来说，简单的机器学习模型主要有两类目的，回归和分类。对于回归而言，我们选择**最小二乘法**，也就是“**平方损失**”（Squared Loss）作为目标函数；对于分类而言，我们选择“**对数几率损失**”（Logistic Loss）。这两种损失和模型是否是线性并没有直接的关系。当然，对于简单模型来说，模型往往是线性的。那么，当模型是线性的，而目标函数又是我们刚才所说的这两类，这种情况下，我们找到的其实就是**线性回归**和**对数几率回归**这两大基本模型。

当我们选择好了目标函数之后，下面一个步骤一般是尝试**根据目标函数寻找参数的最优解**。这一个步骤我们往往需要根据参数**尝试写出参数的梯度**。对于简单的线性模型来说，这一步往往相对比较容易。但是有一些模型，包括深度学习模型，梯度并不是那么直观就能够得到的。这也就直接导致下面的步骤变得更加复杂。

得到梯度以后，一般来说，我们首先**尝试有没有可能得到一个解析解**。

有解析解，往往就意味着我们并不需要通过迭代的方法来得到一个数值优化的解。解析解往往也不是近似解，而是一个确切的答案。当然，在真实的数据中，一些理论上的解析解因为数值稳定性的因素依然无法得到。对于解析解来说，我们需要写出参数的梯度，然后尝试把等式置零，然后看是否能够解出参数的表达式。这个过程并不一定对于每一个模型都适用。

如果我们没法得到解析解，就需要另外一个方法了，那就是利用数值计算来取得一个近似解。在有了梯度信息以后，一种最普遍的数值计算方法就是**梯度下降法**。从原则上来说，梯度下降是求一个函数最小值的数值流程。如果你需要求一个函数最大值的流程，那就需要**梯度上升**。

怎样才能保证梯度下降一定能够得到最优解呢？一般来说，梯度下降并不能保证找到函数参数的最优解，往往只能找到一个局部最优解。对于**凸问题**（Convex Problem）而言，局部最优也就是全局最优。因此，**从理论上说，梯度下降能够找到凸问题的全局最优解**。当然，到底多快能够找到这个最优解，也就是算法的收敛速度是怎样的，就又是另外一个问题了。

但是对于**非凸**（Non Convex Problem）问题而言，梯度下降仅仅能够收敛到一个局部最优解，这个解是否能够被接受，还有待考证。

深度学习模型的优化

在这里，我们从普通的模型衍生出来，看一看深度学习模型的优化问题。

和普通模型一样，**深度学习模型也需要一个目标函数来对参数进行有效学习**。我们前面在介绍深度学习模型的时候提到过，很多时候，深度模型都充当了更加复杂的**特征提取器**的角色。而在最后一层的表达中，我们可以认为是在复杂特征后的线性模型。因此，我们依然可以使用回归（或者说平方损失），抑或分类（或者说是**对数几率损失**），来对不同的问题进行建模。

遗憾的是，深度模型的特点就是多层。而进行优化的流程中，第一个步骤就是梯度的计算，这一步因为模型的多层变得复杂起来。从概念上来说，我们需要得到在当前迭代下，针对模型参数的**梯度**，这些梯度包括每一个隐含层的参数。我们之所以能够计算深度模型，第一个重要发展就是能够计算这些梯度。

在深度学习的语境中，计算梯度的方法叫作“**反向传播**”（Back Propagation）。关于如何计算反向传播，网络上有很多教程，我们在这里就不赘述了。你需要记住的是，**反向传播是为了有效快速地计算梯度**。

那么，有了梯度之后，我们是不是就能够得到深度模型参数的解析解呢？

很可惜，我们无法得到一个解析解。原因是**深度模型的复杂性**以及其**高度的非凸性**。我们不仅无法得到一个解析解，也没有办法轻易得到一个全局最优解。能够采用的一种办法就是使用**梯度下降**来对问题进行**近似求解**。而如何利用梯度下降的办法来对深度模型有效求解，一直都是深度学习研究领域的一个重心。

在过去将近 10 年的研究中，大家发现，一个**好的初始值**往往能够让优化过程变得容易一些。而在梯度下降的过程中，有一些下降方法也要好于其他的方法。一些小技巧，比如 Dropout，批量归一化（Batch Normalization）等，已经变成了深度模型优化的标准流程之一，目的就是能够有效地计算梯度，而且有较好的数值稳定性。

小结

今天我为你讲了机器学习模型，包括传统模型和深度学习模型的优化流程。

一起来回顾下要点：第一，我们聊了什么是模型、目标函数和优化过程这三个概念，清晰理解这三个概念之间的关系非常重要；第二，我们从一般机器学习模型的优化过程入手，讲了深度模型的优化。

最后，给你留一个思考题，有哪些因素造成了深度模型优化的困难？

欢迎你给我留言，和我一起讨论。

AI 技术内参

你的360度人工智能信息助理

洪亮劼

Etsy 数据科学主管
前雅虎研究院资深科学家



新版升级：点击「 请朋友读」，10位好友免费读，邀请订阅更有**现金**奖励。

© 版权归极客邦科技所有，未经许可不得传播售卖。页面已增加防盗追踪，如有侵权极客邦将依法追究其法律责任。

上一篇 144 | 基于深度学习的计算机视觉技术（二）：基本的深度学习模型

下一篇 146 | 计算机视觉领域的深度学习模型（一）：AlexNet

精选留言 (2)

 写留言



吕胜

2018-09-06

洪老师，以后讲解麻烦您加入图片和公式，非常感谢您
展开

 1



Xiangjun

2018-09-09

请问，有没有关于模型优化方面更加深入一点的教程或者材料推荐？
展开



拼课微信：171614366!