

设计、可视化和理解深层神经网络——lec3

一、文献信息

作者: John Canny

题目: Designing, Visualizing and Understanding Deep Neural Networks

发表途径: CS 182/282A

发表时间: Spring 2019

二、问题意义

神经网络算法是最早来源于某神经生理学家和某数学家联合发表的一篇文章,他们对人类神经运行规律的提出了一个猜想,并尝试给出一个建模来模拟人类神经元的运行规律。神经网络一开始由于求解问题的不稳定,以及范围有限被抛弃。后面又在各个科研人员的努力下,对遇到的问题一个个解决,加上因为游戏带来的计算能力的提升获得了一个爆发式的增长。

深层神经网络由于具有数量庞大的参数,使得其具有强大的表征能力。当其神经元的个数足够多时,深度神经网络几乎可以表征任意函数。基于这一点,当我们拥有足够的数据和算力时,由于数据可以弥补参数过多带来的过拟合,算力可以让深度神经网络的训练成为可能,那么我们就拥有了一个对于学习问题的相对通用的方法。

三、思路方法

作者一开始总结上节的内容,生成模型与判别模型的特点: ①深层神经网络的生成模型与判别模型是有区别的模型。②一开始可能学习得比较慢,但渐近精度更好。之后开始进入此次研究的内容: 偏差-方差权衡、正规化、支持向量机(SVM)、多类 SVM、多类 Logistic 和 Softmax、交叉验证。

在偏差-方差均衡研讨中,作者首先推导出期望值:

$$E[f(X)] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(X) P(X) dX_1 \dots dX_n$$

但有时期望被一些东西所取代,而不是我们主要关注的点,在这种情况下,需要指定变量。

接下来进行偏差-方差的分解,对于平方损失问题,总平方误差分解为偏差和方差:

$$E\left[\left(\hat{f}(x) - f(x)\right)^2\right] = E\left[\left(\hat{f}(x) - \bar{f}(x) + \bar{f}(x) - f(x)\right)^2\right]$$

此时需要注意,方差和偏差需要均衡,深度网络具有高维参数化,因此其他模型相比,它们往往具有高方差和低偏差。通过正则化可以减少模型中的方差,这是深度学习的设计。

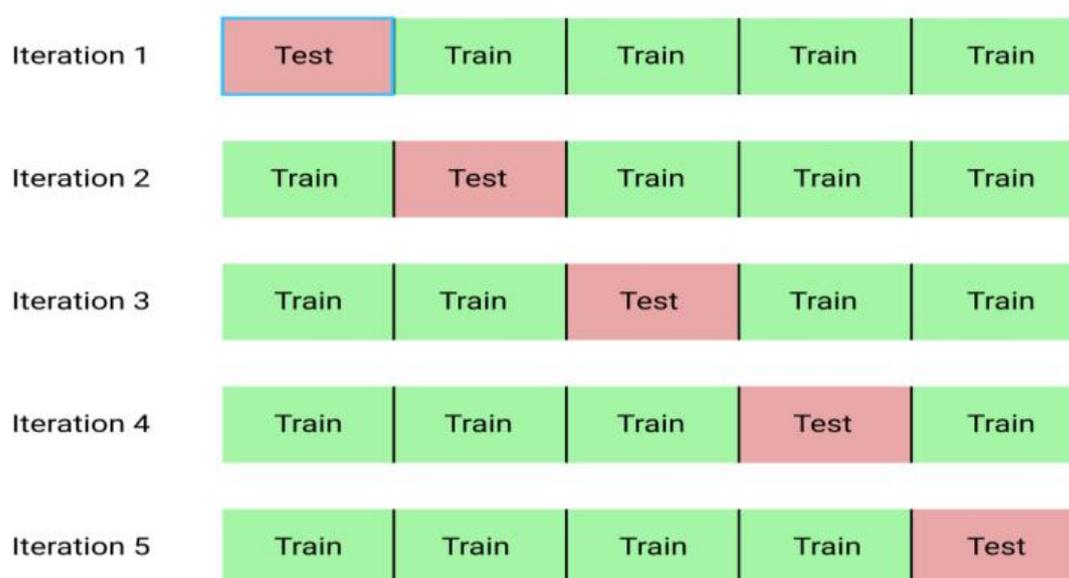
接下来开始讨论规则化的方法,这里我们用到了线性代数。如果 $\mathbf{v}^T \mathbf{M} \mathbf{v} > 0$, 对于所有实向量, 矩阵是正定的, 如果 $\mathbf{v}^T \mathbf{M} \mathbf{v} \geq 0$, 对于所有的实向量, 矩阵是半正定的。如果我们能让 λ 足够大, 那对于我们的规律化将会十分有利。在这里我们用到多元回归的方法, 最后我们可以

得到：强正则化 (λ 大)，可以得到低方差和高偏差，弱正则化 (λ 小)，可以得到高方差和低偏差。

接下来我们对模型的性能进行分析，当我们测量分类器性能时，我们希望使用最精确的模型，所以我们希望使用尽可能多的数据进行训练。但我们也希望尽可能多地用于测试，以便获得精确的测量结果。经过验证，我们可以得到我们可以对数据进行其他分区，如上面的分区。

如果我们保持测试样本不相交，它们将给出大部分独立的测量结果。

我们通过使用 k 倍交叉验证设计（这里 $k=5$ ）来实现这一点。



对于每个迭代，我们将 4 个绿色折叠合并为一个训练集，并训练一个模型。我们在红色褶皱上为迭代计算模型。最后，我们对 5 次迭代的性能进行了平均，可以得到其更好的性能。

四、实验结论

论文使用的论述方法是先提出一个问题再尝试用各种方法来解决、优化它，十分引人入胜。在解决遇到的问题时，用到了概率论、线性代数、通信原理等方面的知识，让我对以往学习的知识有了更加深刻的理解。论文层层论证，最后证明了经过我们的对模型的各种均衡、正规化得到了最佳性能。

五、启发思考

通过对本篇论文的阅读，我学到了很多。对深度神经网络的建模，使用了偏差-方差权衡、正规化、支持向量机 (SVM)、多类 SVM、多类 Logistic 和 Softmax、交叉验证的方法，其中偏差-方差均衡、正规化用到了我们学习到的概率论和线性代数的知识，让我对以往学习的知识有了更加深刻的理解，让我受益匪浅。