

## Deep Learning 读书会第九次讨论记录（由极视角整理）

下面为 2017 年 1 月 15 日 Deep Learning 读书会第七章深度学习的正则化。本次讨论问题由@**短暂愉悦的派大星**整理 整理，并组织讨论。笔记由极视角筱彤整理。如有想加入读书会讨论的，请联系小助手（微信：Extreme-Vision）。

### 讨论话题

- 话题一.** 正则化的定义是：用以修正学习算法——减少泛化误差而不是训练误差的策略。那么只要是减少泛化误差的方法都可称为正则化嘛？（那么像 BN ,或直接给模型加入减少泛化误差的规则 ,也算正则化嘛？或可以说起到了正则化的作用？（短暂愉悦的派大星） ..... 2
- 话题二.** 正则化策略基于正则化估计（通过平衡偏差和方差实现）,那是否意味着正则化策略与模型复杂程度有关？有没有公式直接可以描述正则化与模型复杂程度的关系？（短暂愉悦的派大星） ..... 3
- 话题三.** 在正则化目标函数的过程中需要加入正则化项（或称为惩罚项）,该项前的系数（对应正则的力度系数）除经验取值外还有哪些主流的方法？（短暂愉悦的派大星） ..... 5
- 话题四.** 数据扩增是否会导致过拟合风险？如何控制扩增数据与原始数据的比例，或者有哪些可以注意事项呢（短暂愉悦的派大星） ..... 7
- 写在最后** ..... 11

**话题一. 正则化的定义是 :用以修正学习算法——减少泛化误差而不是训练误差的策略。那么只要是减少泛化误差的方法都可称为正则化嘛？（那么像 BN，或直接给模型加入减少泛化误差的规则，也算正则化嘛？或可以说起到了正则化的作用？（短暂愉悦的派大星）**

**短暂愉悦的派大星**

我的感觉是本章更多是说明正则化的本质，通常的第一反应是防止过拟合，其一个代表不仅是损失函数中的正则化项，更多的是体现正则的作用，即 regularization，英文有“正规化; 规则化”的含义

**极视角小助手**

所以只要起到相关的作用并可定义为正则化？请大家也都出来讨论讨论~没有人吗

**安兴乐**

加快训练速度啊

**九问**

有人，这个问题我觉得基本就这个意思啊

**短暂愉悦的派大星**

嗯，主要是体现正则化的本质。防止过拟合，也是我们一直努力的方向

**极视角小助手**

好的，那我们继续下一个问题吧

**短暂愉悦的派大星**

好

**话题二. 正则化策略基于正则化估计（通过平衡偏差和方差实现），那是否意味着正则化策略与模型复杂程度有关？有没有公式直接可以描述正则化与模型复杂程度的关系？（短暂愉悦的派大星）**

**极视角小助手**

那首先关于这个问题本身大家有什么疑问吗？

**九问**

这个什么意思？没明白

**安兴乐**

正则化并没有和模型复杂度有什么关系吧？

**九问**

我也这么觉得。。

**极视角小助手**

那一般正则化相关的是什么呢

**九问**

主要是数据的分布吧

**极视角小助手**

模型复杂程度又可以说明什么？各位大神可以稍微解释下吗

## 曲晓峰

以二维平面中的曲线拟合为例，这个标准很容易确定，就是最高幂指数。幂指数越多就是拟合的曲线弯折点越多

## 短暂愉悦的派大星

我觉得正则化策略和模型复杂程度不一定有关，这主要取决于正则化策略采取的方式。模型的复杂程度指的是什么？我的理解，一般情况下，比方说一次函数，二次函数，三次函数，他们的模型复杂程度是属于依次递增的，间接的说可以是，模型参数的个数，指的是模型的复杂程度，当然这样描述并不准确，这是我的一种理解。在深度学习中我的理解是网络的深度和宽度，当然与之相对应的是参数的规模大小。

## 九问

嗯，是这样的，一般浅层模型如果不用 BN 这种方法也能收敛，但是像 resnet 这样的深层网络不用 BN 层，完全不能收敛，亲测

## 曲晓峰

不同的正则化方法对模型复杂程度的影响是明显不同的

## 九问

dropout 也不好使

## 极视角小助手

那有关和无关的情况可以概括出来吗？比如什么时候有关？什么时候无关？或者说正则化采取策略的方式是什么样的来决定这个相关度？大家有什么想法吗

## 九问

我感觉没啥特定的规律，都是靠经验，BN 和模型本身没有太大的关系，但是能够加快深层网络的收敛

## 曲晓峰

比如说  $l_2$  范数只是约束了  $w$  的平方和，复杂度并没有降低。而  $l_1$  范数就明确要求减少参数个数

## 短暂愉悦的派大星

因为一说到模型复杂程度，有时候也会想到  $vc$  维，但这个和实际的数据有关系了。或者说，当然这是和本问题不太相关的一点点想法。实际在考虑设计网络大小的时候，也要将数据本身的规模考虑进去。

## 极视角小助手

噢噢~那我们接着第三个问题

**话题三. 在正则化目标函数的过程中需要加入正则化项（或称为惩罚项），该项前的系数（对应正则的力度系数）除经验取值外还有哪些主流的方法？（短暂愉悦的派大星）**

## 短暂愉悦的派大星

我感觉这个问题有点类似于，学习率的选取策略。当然在这个问题里，学习率变成了，正则化项的力度系数。

### **极视角小助手**

那其实这个问题是想确定哪些主流的方法和它们适用的情况吗

### **短暂愉悦的派大星**

对的。

### **极视角小助手**

那首先你觉得有哪些方法呢？各位大神觉得哪些主流方法呢？然后它们对于哪些情况会比较适用呢

### **曲晓峰**

贝叶斯优化吧，典型的超参优化问题

### **短暂愉悦的派大星**

我就知道经验支持一般选取固定的，就类似于在学习率的选择，但是我也查到了其他的一些正则化参数选取方法。比方广义交叉验证等等方法。但是在深度学习模型方面，常见的或者说常用的比较流行的，正则化系数的选择方法有哪些？重点是，在深度学习模型方面，有效快速，效果好。

**极视角小助手**

那曲博士，还有其他推荐的方法吗

**短暂愉悦的派大星**

曲博士，贝叶斯优化，在深度学习方面的超参数优化用的多吗？

**极视角小助手**

另外除了贝叶斯，请问大家还有其他的一些推荐方法吗？要不我们继续下面一个话题吧~

**短暂愉悦的派大星**

嗯嗯

**话题四. 数据扩增是否会导致过拟合风险？如何控制扩增数据与原始数据的比例，或者有哪些可以注意事项呢（短暂愉悦的派大星）**

**Joffery**

数据扩充不就是为了解决过拟合吗

**极视角小助手**

会增加风险吗？

**短暂愉悦的派大星**

这个问题也是我无意当中想到的。通常情况下大家都喜欢问，预训练时的模型跟之后再使用的模型所采用的数据不一样，数据规模大小不一样该采取什么样的处理方法？当然啦，我想的是另外一个角度，比方说，我训练的时候用的是十张图片，我基于这十张图片做一些裁剪，放大缩小，或者是其他变换，产生，1 万张，甚至是 10 万张新图片训练，那么会导致在训练图像上的过拟合吗？

### **极视角小助手**

这个问题是不是也涉及到一个程度的问题

### **短暂愉悦的派大星**

也就是所说的在没有见过的图片的识别上，的效果，怎么样？会不会导致比较差的结果呀！换句话说，我原始生成的图像或者说是数据规模，基于这个原始数据的规模再扩增出的数据，拿来用作训练。会不会起到反效果？当然，我们知道，有正的积极的效果，但有没有反的效果，他们之间的阈值又在哪里？

### **安兴乐**

最后要看的是测试数据集

### **短暂愉悦的派大星**

你说得对。最后要看性能的话还是要拿，没有见过的图片进行测试。也是要看测试数据的情况和，原始数据的情况。才能说明这个问题



**Joffery**

根据以前的实验，比如图像做一次水平翻转，测试集上会提高一两个百分点，效果不是很显著

**短暂愉悦的派大星**

大概知道了，定性的角度来说，确实是，到一定程度，就没有很大效果，或者说没有效果了，但是定量的话还是需要进一步的研究才可以说明。

**Joffery**

不知道 GAN 生成的图像用作数据扩充会不会有效果

**短暂愉悦的派大星**

我刚刚也想到这个问题

**Joffery**

但是目前 GAN 生成的图像效果不是很好吧

**短暂愉悦的派大星**

我看有些小组在做有关比较小的图像的时候，先用生成网络对小图进行放大。然后再做一些检测或者分类的工作。

**九问**

数据扩充要看情况吧，如果样本多样性够的话还行

### 极视角小助手

好的，那我们今晚的问题差不多了。下面还有@短暂愉悦的派大星的一个想法~  
通过观察正则化目标函数，正则化项（或称惩罚项）本身也可看做一个（与网络参数直接相关的）模型。将该等参数（如参数个数，层数，深度等等组合）作为网络输入、原网络最优值时的正则项结果作为目标，训练一个深度网络计算任意深度的最佳正则化项。

### 短暂愉悦的派大星

就是有的不好量化，或者说我目前还没有发现一些比较好的量化，这些图像多样性的手段。当然常见的一些，图像的属性是一个方法。

### 极视角小助手

@短暂愉悦的派大星 可以解释一下最后的想法~

### 短暂愉悦的派大星

这个想法是我想到之前好像看到有一篇文章讲的是用网络可以预测梯度。当然了，我们也可以用网络来预测损失函数中最佳的正则化项的值。

<http://mp.weixin.qq.com/s/vgNRgC19oElwaGO2TzoEVg>

### 极视角小助手

那我们今晚就到这里啦，感谢@短暂愉悦的派大星的问题准备和大家的讨论~如  
还有什么要讨论的，欢迎继续在群里讨论~

---

End

---

## 写在最后

非常感谢此次进行讨论交流的朋友们以及群内支持的朋友们，希望我们读书会能让大家学到更多，并且讨论后可以对原书有更独到的理解。 **如有愿意提交自己对这本书的理解的，欢迎联系小助手，加入此项目组。**

*#广告时间#*

视觉前沿资讯，将算法放至极市关注请关注极市平台公众号。

