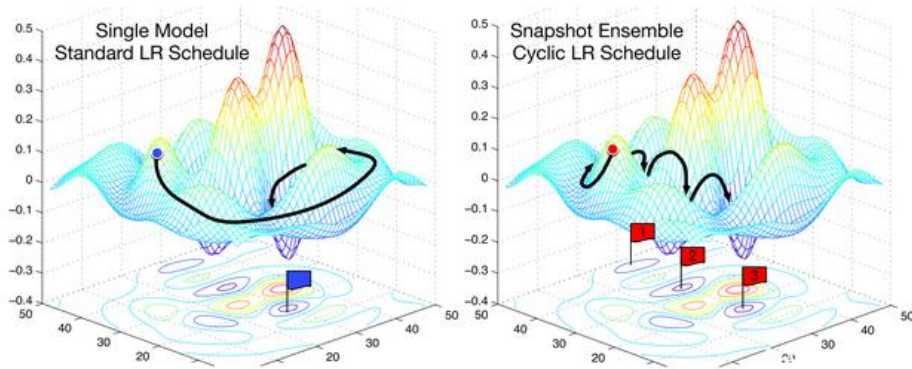


Adam 优化算法详解

[播报文章](#)

deephub

发布时间: 2020-06-05 08:47

[关注](#)

据牛津字典的定义，优化是指最好或最有效地利用一种情况或资源，或者简单地使自己的事物达到最佳状态的行为。通常，如果可以对某事进行数学建模，则很有可能可以对其进行优化。这在深度学习领域起着至关重要的作用（可能是整个人工智能），因为您选择的优化算法可能是在数分钟，数小时或数天（有时甚至是数周）内获得高质量结果的区别。

在这篇文章中，我们将阐述：

- 什么是Adam Optimizer?
- 在深度学习模型中使用Adam进行优化有什么好处?
- Adam如何工作?

什么是Adam Optimizer?

什么是Adam Optimizer是对SGD的扩展，可以代替经典的随机梯度下降法来更有效地更新网络权重。

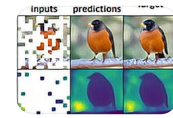
请注意，Adam这个名字并不是首字母缩写词，实际上，作者（OpenAI的Diederik P. Kingma和多伦多大学的Jimmy Lei Ba）在论文中指出，该论文首次在ICLR 2015上作为会议论文发表，标题为Adam: A method for Stochastic Optimization, that the name is derived from adaptive moment estimation.。

作者毫不犹豫地列出了将Adam应用于非凸优化问题的许多迷人好处，我将继续分享以下内容：

简单地实现（我们将在本文的稍后部分中实现Adam，并且您将直接看到如何利用强大的深度学习框架以更少的代码行使实现变得更加简单。）

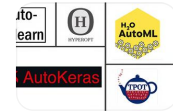
1. 计算效率高
2. 很少的内存需求
3. 梯度的对角线重缩放不变（这意味着亚当将梯度乘以仅带正因子的对角矩阵是不变的，以便更好地理解此堆栈交换）
4. 非常适合数据和/或参数较大的问题
5. 适用于非固定目标

作者最新文章



MultiMAE：一种简单、灵活且有效的...

1天前



自动化的机器学习：5个常用AutoML ...

2天前 10阅读



将梯度提升模型与Prophet 相结合可...

3天前 11阅读

相关推荐



如果训练数据受到损害，机器学习模型...

钲点科技谈



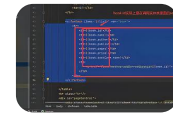
「扩展内容」时间准确度方面，Pytho...

python高手养成 1评论



百信银行高校招聘网申测评专业知识测...

weilingjun104



java查询的数据函数是什么

小高讲科技



Java实现平滑加权轮询算法——降权和...

程序那点事

[换一换](#)

- 1 拜登口误越来越离谱了
- 2 台艺人在沪1天只吃1餐? 国... 热
- 3 160秒混剪中国航天大事记 热
- 4 老坛酸菜面重新上架：在上... 新
- 5 网传为医护跳“谢谢你”致家长... 新
- 6 别被网红印花口罩的颜值骗了
- 7 多地通报“气溶胶传播新冠”
- 8 泽连斯基称抓捕了反对派领导人
- 9 骑手回应“大米不加价不给送”
- 10 浙江一高校现疫情 3千余名师生...



细介绍)

Adam是如何工作的

简而言之，Adam使用动量和自适应学习率来加快收敛速度。

Momentum (动量)

在解释动量时，研究人员和从业人员都喜欢使用比球滚下山坡而向局部极小值更快滚动的类比法，但从本质上讲，我们必须知道的是，动量算法在相关方向上加速了随机梯度下降，如以及抑制振荡。

为了将动量引入我们的神经网络，我们将时间元素添加到过去时间步长的更新向量中，并将其添加到当前更新向量中。这样可以使球的动量增加一定程度。可以用数学表示，如下图所示。

$$V_t = \gamma V_{t-1} + \eta \nabla \theta J(\theta)$$

$$\theta = \theta - V_t$$

动量更新方法，其中 θ 是网络的参数，即权重，偏差或激活值， η 是学习率， J 是我们要优化的目标函数， γ 是常数项，也称为动量。 V_{t-1} （注意 $t-1$ 是下标）是过去的时间步长，而 V_t （注意 t 是下标）是当前时间步长。

动量项 γ 通常被初始化为0.9或类似于Sebastian Ruder的论文《An overview of gradient descent optimization algorithm》中提到的项。

适应性学习率

通过将学习率降低到我们在AdaGrad，RMSprop，Adam和AdaDelta中看到的预定义时间表(schedule)，可以将自适应学习率视为训练阶段的学习率调整。这也称为学习率时间表 有关该主题的更多详细信息，Suki Lau撰写了一篇有关该主题的非常实用的博客文章，称为“Learning Rate Schedules and Adaptive Learning Rate Methods for Deep Learning.”。

在不花太多时间介绍AdaGrad优化算法的情况下，这里将解释RMSprop及其在AdaGrad上的改进以及如何随时间改变学习率。

RMSprop（即均方根传播）是由Geoff Hinton开发的，如《An Overview of Gradient Descent Optimization Algorithms》所述，其目的是解决AdaGrad的学习率急剧下降的问题。简而言之，RMSprop更改学习速率的速度比AdaGrad慢，但是RMSprop仍可从AdaGrad（更快的收敛速度）中受益-数学表达式请参见下图

$$E[g^2]_t = 0.9E[g^2]_{t-1} + 0.1g_t^2$$

$$\Theta_{t+1} = \Theta_t - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}} \cdot g_t$$

$E[g^2]_t$ 的第一个方程是平方梯度的指数衰减平均值。Geoff Hinton建议将 γ 设置为0.9，而学习率 η 的默认值为0.001

这可以使学习率随着时间的流逝而适应，这很重要，因为这种现象也存在于Adam中。当我们将两者（Momentum 和RMSprop）放在一起时，我们得到了Adam —下图显示了详细的算法。

Require: α : Stepsize

Require: $\beta_1, \beta_2 \in [0, 1]$: Exponential decay rates for the moment estimates

Require: $f(\theta)$: Stochastic objective function with parameters θ

Require: θ_0 : Initial parameter vector

$m_0 \leftarrow 0$ (Initialize 1st moment vector)

$v_0 \leftarrow 0$ (Initialize 2nd moment vector)

$t \leftarrow 0$ (Initialize timestep)

```
m_t ← β_1 · m_{t-1} + (1 - β_1) · g_t (Update biased first moment estimate)
v_t ← β_2 · v_{t-1} + (1 - β_2) · g_t^2 (Update biased second raw moment estimate)
m_hat_t ← m_t / (1 - β_1^t) (Compute bias-corrected first moment estimate)
v_hat_t ← v_t / (1 - β_2^t) (Compute bias-corrected second raw moment estimate)
θ_t ← θ_{t-1} - α · m_hat_t / (sqrt(v_hat_t) + ε) (Update parameters)
end while
return θ_t (Resulting parameters)
```

如果你听过吴恩达老师的深度学习课程，吴恩达老师说过“Adam可以理解为加了Momentum的 RMSprop”，上图的公式就是吴恩达老师这句话的由来。

参考资料

- 1. Kingma, D and Ba, J . (2015) Adam: A method for Stochastic Optimization.
- 2. Ruder, S. (2017) An overview of gradient descent optimization algorithms.
- 3. Duchi, J et al. (2011) Adaptive Subgradient Methods for Online Learning and Stochastic Optimization.

[举报/反馈](#)

发表评论

发表神评妙论

发表

评论列表（5条）

- 

你说我没情调

adam优化器可以减少内存的使用

2021-05-05

回复 赞
- 

llllfdsa

一倍? 那不就是一样多吗

01-05 13:28

回复 赞
- 

JiaMing5139

减少? adam使用的参数是adagrad的一倍

2021-08-03

回复 3
- 

俱兰英f1

异常清晰，十分感谢

2021-08-24

回复 1
- 

百度网友34ded7319

简明扼要，好!

02-22 18:47

回复 赞

没有更多啦



[设为首页](#) © Baidu [使用百度前必读](#) [意见反馈](#) [京ICP证030173号](#)
[京公网安备11000002000001号](#)