**深度学习中的Epoch，Batchsize，Iterations，都是什么鬼？**

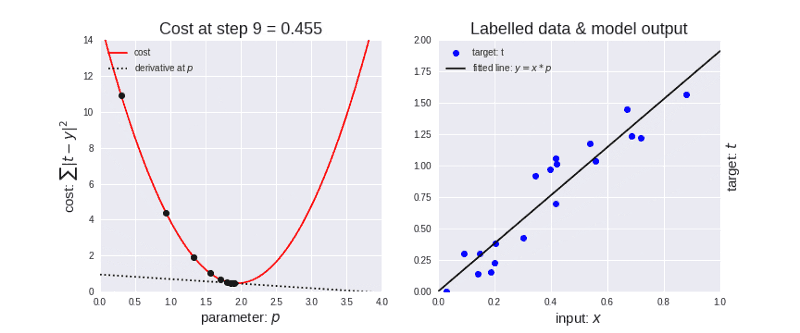
转载自：公众号“习悦智能”，原文链接：https://mp.weixin.qq.com/s/tEOvxPRPAPndNk3thT8EZA

**梯度下降法**

一切的一切，要从机器学习中的梯度下降法说起。

首先让我们来回顾一下这个常见的不能再常见的算法。梯度下降法是机器学习中经典的优化算法之一，用于寻求一个曲线的最小值。所谓"梯度"，即一条曲线的坡度或倾斜率，"下降"指代下降递减的过程。

梯度下降法是迭代的，也就是说我们需要多次计算结果，最终求得最优解。梯度下降的迭代质量有助于使输出结果尽可能拟合训练数据。



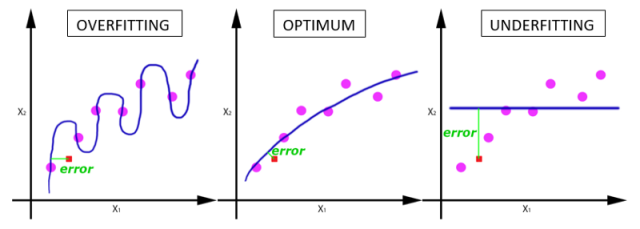
梯度下降法中有一个称为学习率的参数，如上图左所示，在算法开始时，步长更大，即学习率更高。随着点的下降，步长变短，即学习率变短。此外，误差函数也在不断减小。

在训练模型时，如果训练数据过多，无法一次性将所有数据送入计算，那么我们就会遇到epoch，batchsize，iterations这些概念。为了克服数据量多的问题，我们会选择将数据分成几个部分，即batch，进行训练，从而使得每个批次的数据量是可以负载的。将这些batch的数据逐一送入计算训练，更新神经网络的权值，使得网络收敛。

**Epoch**

一个epoch指代所有的数据送入网络中完成一次前向计算及反向传播的过程。由于一个epoch常常太大，计算机无法负荷，我们会将它分成几个较小的batches。那么，为什么我们需要多个epoch呢？我们都知道，在训练时，将所有数据迭代训练一次是不够的，需要反复多次才能拟合收敛。在实际训练时，我们将所有数据分成几个batch，每次送入一部分数据，梯度下降本身就是一个迭代过程，所以单个epoch更新权重是不够的。

下图展示了使用不同个数epoch训练导致的结果。



可见，随着epoch数量的增加，神经网络中权重更新迭代的次数增多，曲线从最开始的不拟合状态，慢慢进入优化拟合状态，最终进入过拟合。

因此，epoch的个数是非常重要的。那么究竟设置为多少才合适呢？恐怕没有一个确切的答案。对于不同的数据库来说，epoch数量是不同的。但是，epoch大小与数据集的多样化程度有关，多样化程度越强，epoch应该越大。

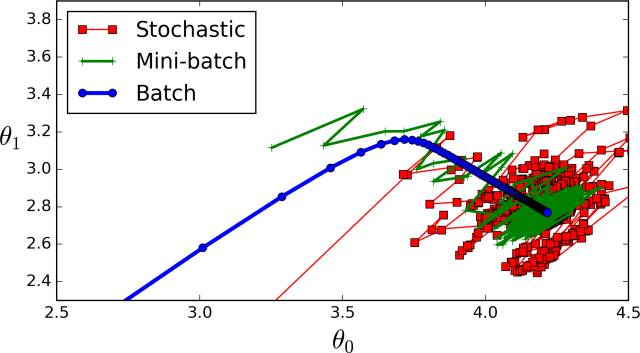
**Batch Size**

所谓Batch就是每次送入网络中训练的一部分数据，而Batch Size就是每个batch中训练样本的数量

上文提及，每次送入训练的不是所有数据而是一小部分数据，另外，batch size 和batch numbers不是同一个概念~

Batch size大小的选择也至关重要。为了在内存效率和内存容量之间寻求最佳平衡，batch size应该精心设置，从而最优化网络模型的性能及速度。

下图为batch size不同数据带来的训练结果，其中，蓝色为所有数据一并送入训练，也就是只有1个batch，batch内包含所有训练样本。绿色为minibatch，即将所有数据分为若干个batch，每个batch内包含一小部分训练样本。红色为随机训练，即每个batch内只有1个训练样本。



上图可见，蓝色全数据效果更好，当数据量较小，计算机可以承载的时候可以采取这种训练方式。绿色的mini分批次训练精度略有损失，而红色的随机训练，难以达到收敛状态。

**Iterations**

所谓iterations就是完成一次epoch所需的batch个数。

刚刚提到的，batch numbers就是iterations。

简单一句话说就是，我们有2000个数据，分成4个batch，那么batch size就是500。运行所有的数据进行训练，完成1个epoch，需要进行4次iterations。

怎么样，你成功被绕晕了吗？

阅读 105评论 0赞 0