神经网络基础————Batch Normalization

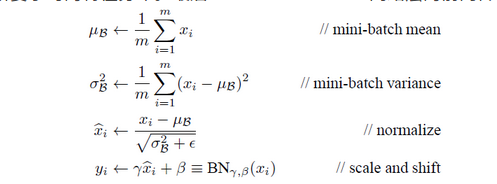
在神经网络中出现了一个很火的处理方法，称为Batch Normalization。与卷积层、池化层一样BN层也是神经网络设计中的层。从表面上来看，BN是将输入数据进行归一化为均值是0，方差是1的标准分布。深入分析，该方法有精妙之处。

一.BN的目的

在神经网络训练时，数据的输入是分批次输入的，所谓的Batch便是这个概念。而每个批次的数据虽然来自训练数据集，但是数据的分布之间总会有相应的差异。分批次输入的目的是提高网络的泛化能力，但是正是数据分布的差异使得网络的训练速度大大减慢。例如：第二批输入数据与第一批输入数据分布是有差异的，内部的参数便要通过训练再尽量拟合第二批数据的分布，这种变化被称为：ICS（内部变量转移）。故在神经网络训练的过程中就一直在适应不同的网络数据分布，使得训练速度大幅度减慢。造成这种问题的主要原因是训练数据与测试数据分布不一样，我们不能去尽量训练一个网络，使其十分满足训练数据的分布，这样网络的泛化能力会变差。同时我们也只能根据训练数据训练。故有一个思路便是尽量让输入数据的Batch之间的数据分布相似，这样便可以适当避免ICS。

二.BN的实现过程

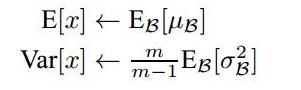
前向传播：



1.求均值和方差。对于数据数据Batch为B，channel为C，尺寸为W，H。则其均值和方差是分通道的。即是对B×W×H这么多数据求均值和方差，共有C组均值和方差。（因为在后续的训练中，不同通道之间的数据是有求和运算的，故不对跨通道）。

2.如果对于数据单纯的求均值和方差，那么其本身的数据分布必然发生很大的变化，前一层所学得的特征便会被完全的打破，相当于后面一层破坏前面一层，特征提取会有很大的影响。为了保存前面学习到的数据分布，设计者脑洞大开，设计了需要学习的变量beta和gama。这两个变量是需要学习的，更新的过程是以Loss为基础的反向传播。beta和gama也采用参数共享的原则。即每一个W×H共享一对值，共有B×C对参数。

3.测试阶段，由于网络的参数已经固定，而输入的测试图像的数量往往是某一个或某几个。故在测试阶段均值和方差应用的是在训练阶段每个Batch的均值和方差的期望：



三.BN的实质

个人理解：BN的实质是提高数据之间（Batch与Batch、训练数据与测试数据之间）的相似性，通过不断的调整beta和gama的目的是为了提高数据分布的相似性。

还有人会从反向传播的角度来解释，这个方法之后会研究。