# 4. Batch Normolization

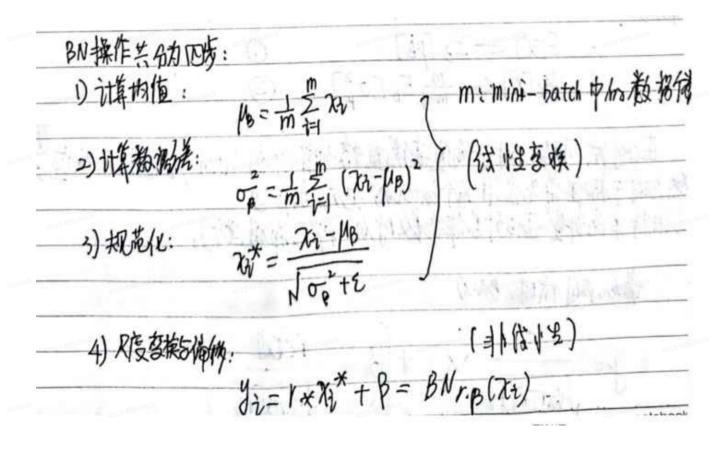
## 1.引入

19. 深度等习主Batch Normalization 抓作用一种经网络在训练的时候随着网络最新加强,激为函数的输入值的整体分布逐渐性激出函数的取值区间上下限靠近,从不在处行传播时代最初神经网络的梯度消失。(私入在其本不多,核场》)

而Butch Normalization 纳作用是通过关现范化物分段,将故来或偏的分布过回到标准似的分布,使将激为函数的输入值存在激步函数对输入物转敏感的区域,从而使标度变大,加快等习物效速度,避免标度消失的问题

# 2.原理

BN(Batch Normolization)是Google提出的用于解决深度网络**梯度消失和梯度爆炸**的问题,可以起到一定的正则化作用。我们来说一下它的原理:



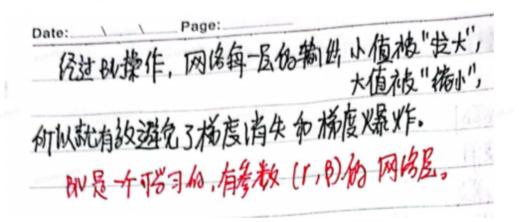
前3步就是对批规范化,使得结果(各个维度)的均值为0,方差为1。

最后一步目标: **让我们的网络可以学习恢复出原始网络所要学习的特征分布** 

#### 1.1BN的gama labada意义

对网络某一层**A**的输出数据做**归一化**,然后送入网络下一层**B**,这样是会**影响到本层网络A所学习到的特征的**。于是**BN**最后的"**尺度变换和偏移**"操作。

引入了这个可学习重构参数γ、β,**让我们的网络可以学习恢复出原始网络所要学习的特征分布**。



## 3.BN训练和测试有什么不同

训练时,<mark>均值和方差针对一个Batch</mark>。

测试时,<mark>均值和方差针对**整个数据集**而言</mark>。因此,在训练过程中除了正常的前向传播和反向求导之外,我们还要记录**每一个Batch的均值和方差**,以便训练完成之后按照下式计算**整体的均值和方差**。

测试模型中,对于均值来说直接计算所有batch  $\mu\beta$ 值的平均值;然后对于标准偏差采用每个batch  $\sigma\beta$ 的无偏估计(无偏估计是用样本统计量来估计总体参数时的一种无偏推断)。

$$\mathrm{E}[x] \leftarrow \mathrm{E}_{\mathcal{B}}[\mu_{\mathcal{B}}]$$
 $\mathrm{Var}[x] \leftarrow \frac{m}{m-1} \mathrm{E}_{\mathcal{B}}[\sigma_{\mathcal{B}}^2]$ 

最后测试阶段,BN的使用公式就是:

$$y = \frac{\gamma}{\sqrt{\operatorname{Var}[x] + \epsilon}} \cdot x + \left(\beta - \frac{\gamma \operatorname{E}[x]}{\sqrt{\operatorname{Var}[x] + \epsilon}}\right)$$

$$+ \operatorname{FR} \circ \operatorname{FD} \circ \operatorname{FR}$$

### 4.LN与BN

LN:Layer Normalization,LN是"横"着来的,<mark>对一个样本,经过同一层的所有神经元做归一化</mark>。 LN中同层神经元输入拥有相同的均值和方差,不同的输入样本有不同的均值和方差;

BN: Batch Normalization,BN是"竖"着来的,<mark>经过一个神经元的所有样本做归一化</mark>,所以与**batch size**有关系。

BN中则针对不同神经元输入计算均值和方差,同一个batch中的输入拥有相同的均值和方差。

#### Z = WX

LN是对W的一个限制,需要它进行归一化处理 BN是对X的一个限制,让其进行归一化处理

二者提出的目的都是为了加快模型收敛,减少训练时间。

