

Normalization

1 Normalization

在神经网络训练的过程中，会出现 ICS(Internal Covariate Shift, 内部协变量漂移) 的问题，表现为：

(1) 在训练时，每层网络输出的分布都在变化，使深层网络需不断适应浅层分布变化，从而降低了网络收敛速度；

(2) 在训练时，模型容易陷入激活函数的饱和区 (如 Sigmoid 和 Tanh)，使梯度变小，从而降低了参数更新速度，甚至使学习过早停止；

(3) 在训练时，每层更新都会影响到其它层，因此每层的参数更新策略需尽可能谨慎。

为了处理 ICS 问题，提出了 Normalization 方法。Normalization 的主要思想是对每一层的输入进行归一化处理，使得每一层的输入分布保持稳定，从而加速训练过程。

2 不同 Normalization 方法

$$\text{Normalization}(x) = \frac{x - \mathbb{E}[x]}{\sqrt{\text{Var}(x) + \epsilon}} * \gamma + \beta \quad (1)$$

2.1 Batch Normalization

特点：Batch Normalization 是对一个 batch size 样本内的每个特征做归一化，即所有样本在同一特征维度上计算均值和方差。

2.2 Layer Normalization

特点：Layer Normalization 是对一个样本内的所有特征做归一化，即每个样本在所有特征维度上计算均值和方差。

2.3 RMS Normalization

特点：RMS Normalization 是对一个样本内的所有特征做归一化，和 Layer Normalization 相似；不同之处在于

（1）RMS Normalization 使用均方根（Root Mean Square）来计算标准差，而不是使用方差。

（2）RMS Normalization 舍弃中心化（即不减去均值），只进行缩放。

3 参考

Batch Normalization

Layer Normalization

RMS Normalization

详解 Normalization