Bach Normalization

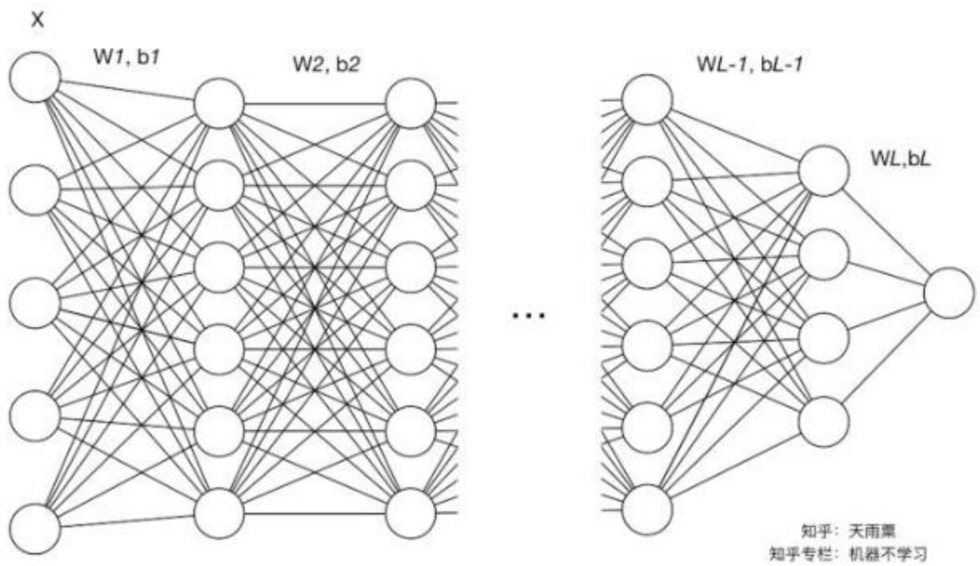
- 算法提出背景
- BN思路与原理
- BN公式
- 可视化BN过程

• 算法提出背景

1. 神经网络中的“**Internal Covariate Shift**”问题

• 定义

在深层网络训练的过程中，由于网络中参数变化而引起内部结点数据分布发生变化的这一过程被称作Internal Covariate Shift。



每一层的线性变换：$Z^{[l]} = W^{[l]} \times input + b^{[l]}$

非线性变换：$A^{[l]} = g^{[l]}(Z^{[l]})$

在每一层W和b都会更新，Z的值相应发生变化，其分布也会改变，A也会随着改变，这一过程称为Internal Covariate Shift

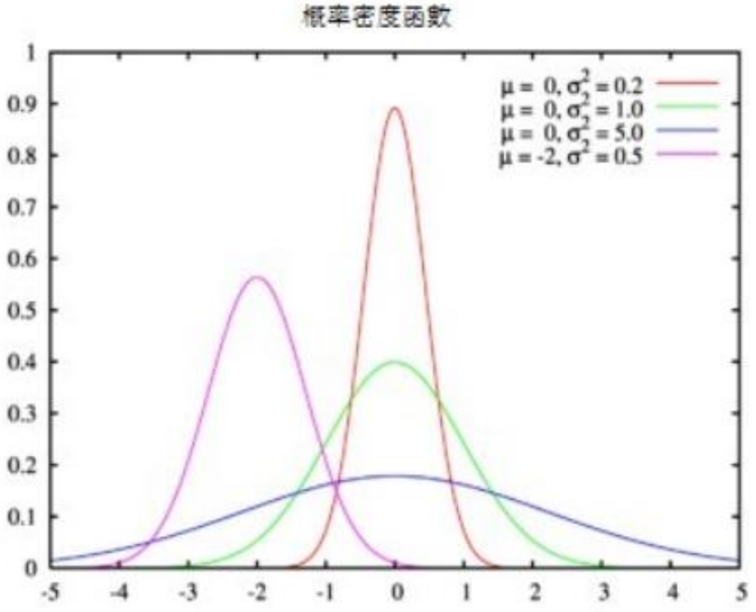
2. **Internal Covariate Shift带来的问题**

（1）上层网络需要不停调整来适应输入数据分布的变化，导致网络学习速度的降低

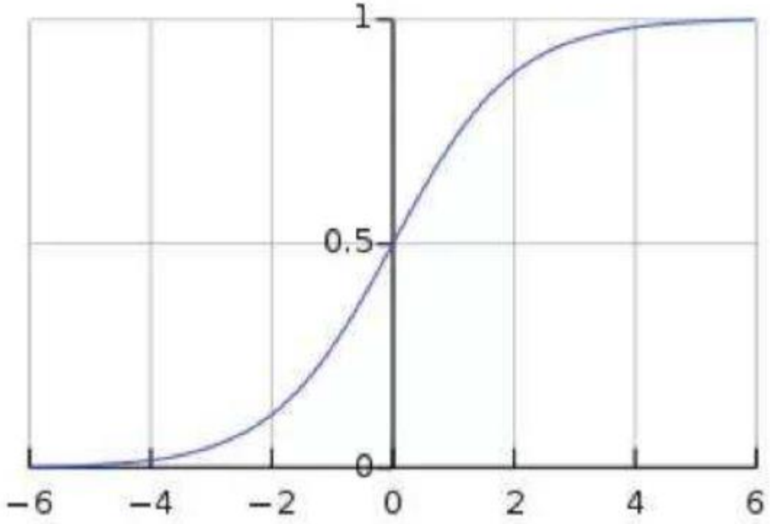
每一层的W和b改变 $\longrightarrow$ 分布改变 $\longrightarrow$ 适应这一改变导致学习率降低。

（2）网络的训练过程容易陷入梯度饱和区，减缓网络收敛速度

正太分布图:



Sigmoid函数:



- Bach Normalization

1. 思路

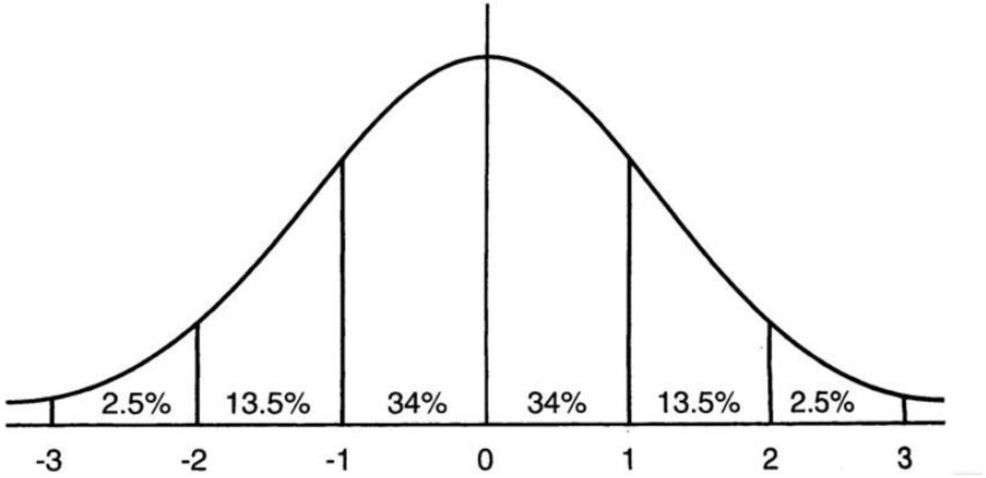
（1）对每一层进行标准化操作

（2）再增加一次线性变换，尽可能恢复数据本身的表达能力

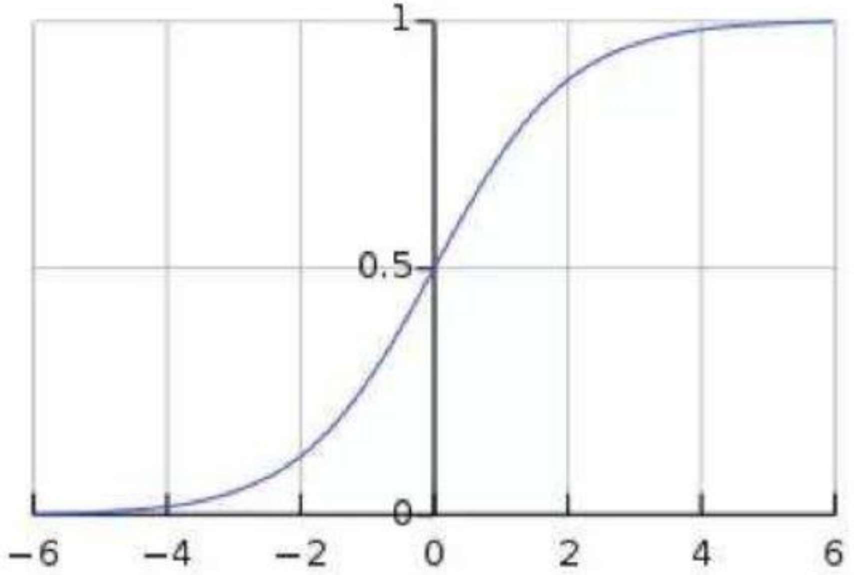
2. **对每一层进行标准化操作**

将每一层的分布转化为均值为0，方差为1的标准正太分布。

标准正太分布：

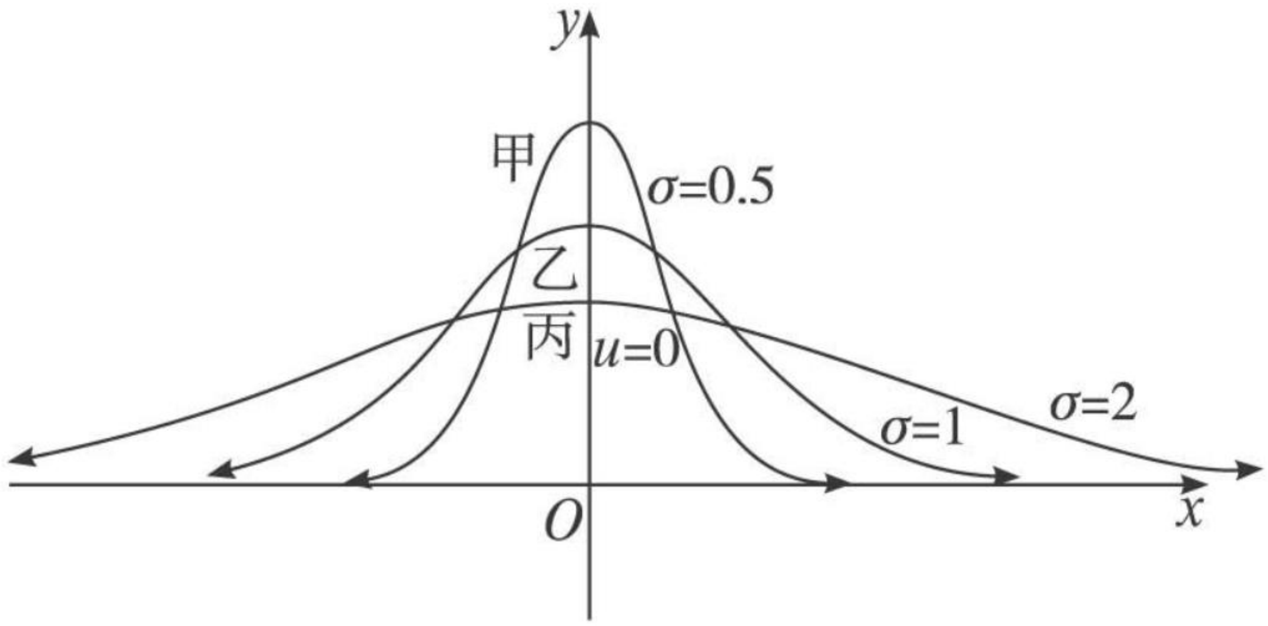


Sigmiod激活函数：



3. **增加一次线性变换**

- **BN为了保证非线性的获得，对变换后的满足均值为0方差为1的Z又进行了scale加上shift操作(Y=scale*Z+shift)**
- **每个神经元增加了两个参数scale和shift参数，这两个参数是通过训练学习到的**



- Bach Normalization公式

$$Z^{[l]} = W^{[l]}A^{[l-1]} + b^{[l]}$$

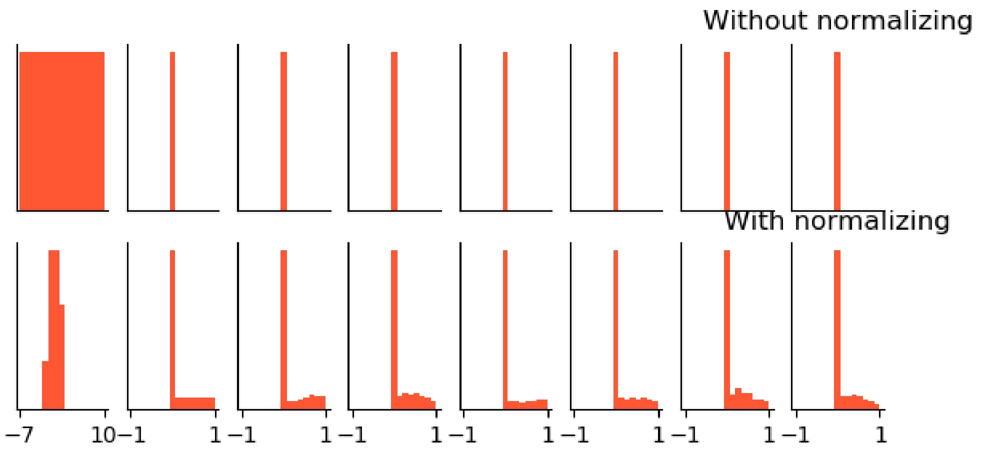
$$\mu = \frac{1}{m}\sum_{i=1}^{m} Z^{[l](i)}$$

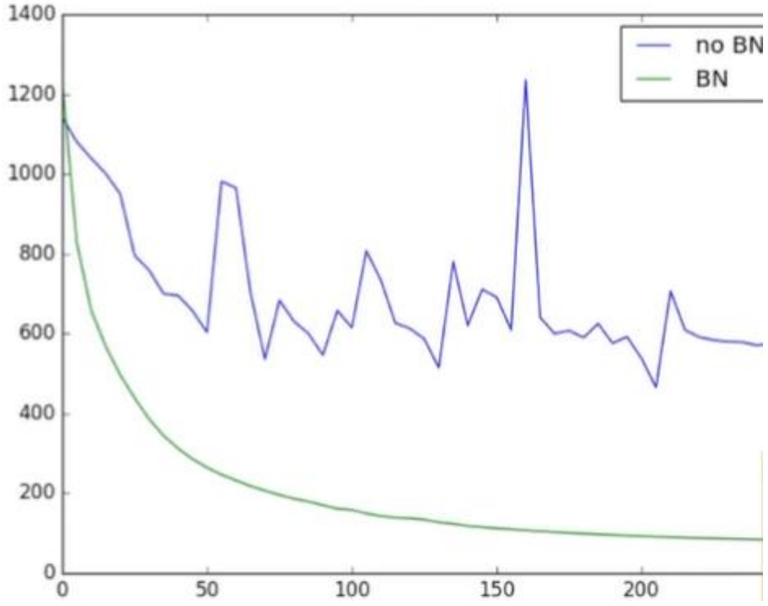
$$\sigma^2 = \frac{1}{m}\sum_{i=1}^{m}(Z^{[l](i)} - \mu)^2$$

$$\tilde{Z}^{[l]} = \gamma \cdot \frac{Z^{[l]} - \mu}{\sqrt{\sigma^2 + \epsilon}} + \beta$$

$$A^{[l]} = g^{[l]}(\tilde{Z}^{[l]})$$

- 可视化BN过程


Without normalizing
With normalizing
−7
10−1
1 −1
1 −1
1 −1
1 −1
1 −1
1 −1
1


no BN
BN
1400
1200
1000
800
600
400
200
0
0
50
100
150
200