

(2011a) 建议使用标准初始化 (normalized initialization)

$$W_{i,j} \sim U \left(-\sqrt{\frac{6}{m+n}}, \sqrt{\frac{6}{m+n}} \right). \quad (8.23)$$

这后一种启发式方法初始化所有的层，折衷于使其具有相同激励方差和使其具有相同梯度方差之间。该式由不含非线性的链式矩阵乘法网络的假设推导得出。现实的神



