

例如，假设某个计算图中包含一条反复与矩阵 W 相乘的路径。那么 t 步后，相当于和 W^t 相乘。假设 W 有特征值分解 $W = V \text{diag}(\lambda) V^{-1}$ 。在这种情况下，很容易看到

$$W^t = (V \text{diag}(\lambda) V^{-1})^t = V \text{diag}(\lambda)^t V^{-1}. \quad (8.11)$$

当特征值 λ_i 不在 1 附近时，若在量级上大于 1 则会爆炸；若小于 1 时则会消失。梯度消失与爆炸问题 (vanishing and exploding gradient problem) 是指该计算图上的梯度也会因为 $\text{diag}(\lambda)^t$ 大幅度变化。梯度消失使得难以知道参数朝哪个方向移动能



