

不幸的是，这些初始权重的最佳准则往往不会带来最佳效果。这可能有三种不同的原因。<sup>3</sup>首先，①我们可能使用了错误的标准——它实际上并不能有利于保持整个网络信号的范数。其次，②初始化时强加的性质可能在学习开始进行后不能保持。最后，③该标准可能成功提高了优化速度，但意外地增大了泛化误差。在实践中，我们



