

训练算法中鞍点扩散会有哪些影响呢？^①对于只使用梯度信息的一阶优化算法而言，情况是不明的。^②鞍点附近的梯度通常会非常小。另一方面^③实验中梯度下降似乎可以在许多情况下逃离鞍点。Goodfellow *et al.* (2015) 可视化了最新神经网络的几个



