训练算法中鞍点扩散会有哪些影响呢?对于只使用梯度信息的一阶优化算法而言,情况是不明的。鞍点附近的梯度通常会非常小。另一方面②实验中梯度下降似乎	
言、情况是不明的 0 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
	·



