

Deep Learning(Ian Goodfellow, Yoshua Bengio, Aaron Courville) 学 习笔记

李奥林

mr liaolin@outlook.com

目录

第一章 数值计算	1	1.3 基于梯度的优化方法	1
1.1 上溢和下溢	1	1.3.1 梯度之上：雅克比和海森矩阵	1
1.2 病态条件数	1		

第一章 数值计算

1.1 上溢和下溢

1.2 病态条件数

1.3 基于梯度的优化方法

1.3.1 梯度之上：雅克比和海森矩阵

二阶推导的前提是要最小化的函数能用二次函数很好地近似，此时海森的特征值决定了学习速率的量级。

函数在特定方向 \mathbf{d} 上的二阶导数可以写成 $\mathbf{d}^T \mathbf{H} \mathbf{d}$ 。当海森矩阵是正定的（所有特征值都是正的），则对于任意方向，二阶导数均为正值，则临界点为局部最小点。当所有非零特征值是同号的且至少有一个特征值是 0 时，这个检测就是不确定的，这是因为单变量的二阶导数为 0 的点的二阶导数测试是不确定的。