P52

第四章 数值计算

优化是指改变x大小以最小化或最大化某个函数f的任务。

导数是在x点处的斜率，换句话说，它表明如何缩放输入的小变化才能在输出获得相应的变化。因此导数对于最小化一个函数很有用，因为它告诉我们如何更改x来略微改善y。

沿着梯度相反的方向进行下降就是梯度下降

反向传播的梯度计算

不是采用函数定义的数值求导，而是反向传播中的批量求导

需要计算输入和输出都为向量的函数的所有偏导数，包含所有这样的偏导数的矩阵被称为Jacobian矩阵。

二阶导数确定函数的曲率，对于负曲率，代价函数实际上比梯度预测下降的更快，没有曲线时，梯度正确预测下降值。对于正曲率，代价函数比预期下降得更慢，并且最终会开始增加，因此太大的步骤，实际上可能会无意的增加函数值。

当我们的函数具有多维输入时，二阶导数也有很多，我们将这些导数合并成一个矩阵，称为Hessian。可以通过Hessian的特征值判断函数的点是不是全局最优点。