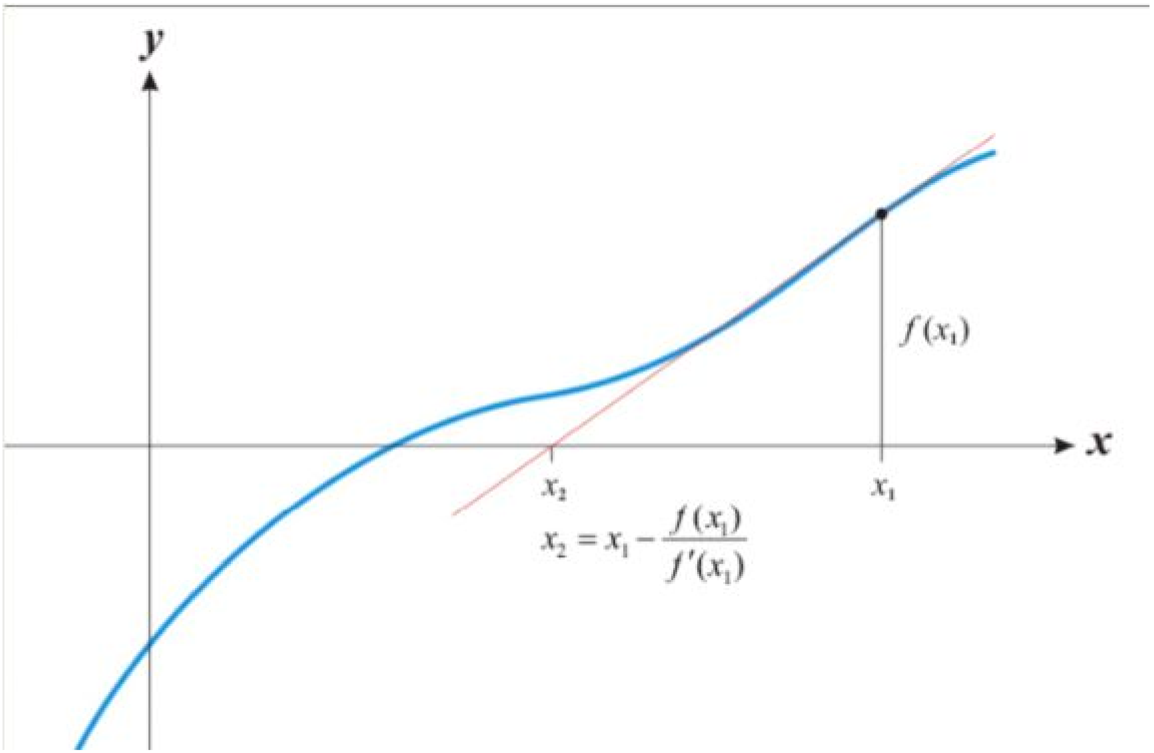
牛顿法、拟牛顿法（BFGS），L-BFGS算法

人在旅途
一起学习人工智能

1 什么是牛顿法？

牛顿法最初是为了求解方程的根而推导出来的公式。



具体步骤：

解：

$$(x-x_1)f'(x_1)=y-f(x_1) \qquad (1)$$

令上式中 $y=0$ ，求得：

$$x=x_1-\frac{f(x_1)}{f'(x_1)} \qquad (2)$$

此时，将新求解的点 x 的坐标命名为 x_2 ，通常 x_2 是比 x_1 更接近方程 $f(x)=0$ 的解。利用 x_2 开始进行下一轮的迭代。迭代公式可以简化如下：

$$x_{n+1}=x_n-\frac{f(x_n)}{f'(x_n)} \qquad (3)$$

当迭代结果与上一次的迭代结果相同或者小于一定阈值时，本次的结果即为函数的根。

由于牛顿法是基于当前位置的切线来确定下一次的位置，所以牛顿法又被很形象地称为是"切线法"。**以上即为牛顿法的相关推导。**

2 利用牛顿法求驻点（也即求解函数的解）

当函数 $f(x)$ 的一阶导数 $f'(x)$ 时点（x, $f'(x)$ ）为函数的驻点。

求某函数的驻点，即为求解该函数的导函数的根，故同样可以利用牛顿法求解。

$$x_k = x_{k-1} - \frac{f'(x_k)}{f''(x_{k-1})} \qquad (4)$$

其中，式(7)中分母是二阶导（此处有点模糊）。

3 牛顿法求驻点的本质

任意一点在xk附近的二阶泰勒展开公式为：

$$f(x) = f(x_k) + f'(x_k)(x - x_k) + \frac{1}{2} f'(x_k)(x - x_k)^2 \qquad (5)$$

当 $f(x) = 0$ 时：

$$f'(x_k) + f''(x_k)(x - x_k) = 0 \qquad (6)$$

由式(6)得：

$$x_k = x_{k-1} - \frac{f'(x_k)}{f''(x_{k-1})} \qquad (7)$$

其中，式(7)中分母是二阶导（此处有点模糊）。

由此，我们可以得出结论：牛顿法的本质是泰勒级数的二阶展开。

4 多元函数的求导问题

一阶导数 f'(x) → 梯度 $\nabla f = \begin{bmatrix} \vdots \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_N} \end{bmatrix}$

二阶导数 f''(x) → Hessian 矩阵

$$\nabla^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_N} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_N} \\ & & \ddots & \\ \frac{\partial^2 f}{\partial x_N \partial x_1} & \frac{\partial^2 f}{\partial x_N \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_N^2} \end{bmatrix}_{N \times N},$$

5 什么是BFGS算法？

首先BFGS算法的由来是四位国外大牛的名字缩写组成。该算法提出于1970年。

5.1 BFGS算法产生原因？

式(7)可以表示为：

$$X_{k+1} = X_k - H_k^{-1} \cdot g_k, \quad k=0,1,\cdots \tag{8}$$

由于H矩阵维度超大，求逆矩阵非常困难，计算复杂度非常高。

5.2 BFGS算法过程：

k

逼近方法：

$$D_{k+1}=\left(I-\frac{s_k y_k^T}{y_k^T s_k}\right)D_k\left(I-\frac{y_k s_k^T}{y_k^T s_k}\right)+\frac{s_k s_k^T}{y_k^T s_k} \tag{9}$$

其中，I为单位矩阵。

迭代过程如下所示：

1、其中，I为单位矩阵，D_1初始值也设置为单位矩阵I
2、首先随机一个$\bar{x}_1$，那么损失函数在$\bar{x}_1$处的梯度g_1便可计算得到
3、$\bar{x}_2=\bar{x}_1-D_1\cdot g_1$，通过此式可得$\bar{x}_2$，故又可求得$g_2$
4、$s_1=\bar{x}_2-\bar{x}_1$，通过此式可得s_1
5、$y_1=g_2-g_1$，通过此式可得y_1 (10)
6、将以上结果带入公式(9)可得D_2
7、有D_2可得$\bar{x}_3$，…，然后进行迭代
8、当D_n达到一定的阈值时，停止迭代，
此时可近似看作H_k^{-1}的值。

6 L-BFGS算法（L指的是limited memory）

由于BFGS算法每次都要存储 D_k 矩阵，故需要很大的内存空间对其进行存储，因而可能使得计算机无法正常运行。

以上即为L-BFGS算法。

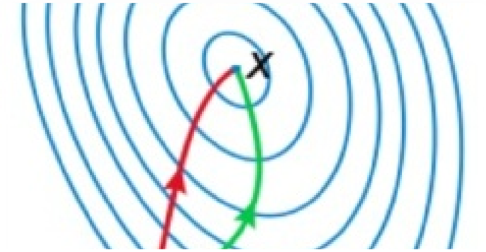
编辑于 2018-09-25

科技　数学　算法（书籍）

文章被以下专栏收录

Daniel的机器学习算法探索道路
人工智能当下是一个流行且热门的行业，它并不是一种盲目跟风，它可以在很多领域...
进入专栏

推荐阅读



梯度下降法、牛顿法和拟牛顿法

牛顿法和拟牛顿法

导言牛顿法和拟牛顿法也是求解无约束最优化问题的常用方法【另一常用方法为：梯度下降法】，有收敛速度快的优点。牛顿法是迭代算法，每一步需要求解目标函数的Hessian 矩阵的逆矩阵，计算...

无法“解”出的方程——一个关于五次方程没有求根公式的...

古典代数学有一个经典的结论，五次（含）以上的方程没有求根公式，这里所谓的求根公式只涉及加、减、乘、除以及开任意次根；这一结论被称为Abel-Ruffini定理。要得到上述结论通常需要借助抽...

牛顿法和拟牛顿法——（书中附录B）

牛顿法(Newton method)和拟牛顿法(quasi-Newton method)也是求解无约束最优化问题的常用方法，具有收敛速度快的优点。牛顿法是迭代算法，每一步需要求解目标函数的海赛矩阵的逆矩阵，计算比...

5 条评论

切换为时间排序

写下你的评论...

有志青年哈哈哈 11 个月前

（5）-（6）条件不对吧。。这推导的太草率了。。

2

顾周 9 个月前

请问（5）-（6）的条件是咋回事？（6）是怎么过来的？看不懂呀。。

赞

霄霄荔枝 回复 顾周 8 个月前

应该是求导等于零，因为是驻点，函数在该点导数应该为零，进而得到6式。

赞

人在旅途 (作者) 回复 霄霄荔枝 8 个月前

好的，多谢提出问题，我稍后及时纠正！

赞

展开其他 1 条回复