原 牛顿法与Hessian矩阵

2017年03月03日 22:13:30

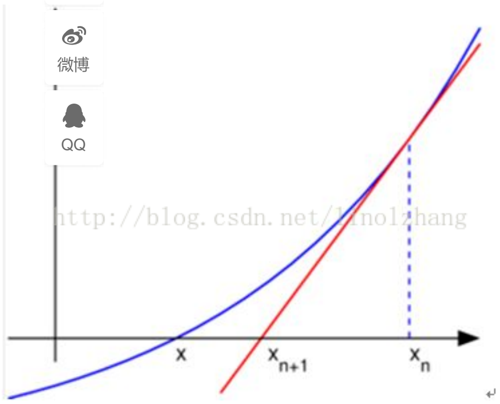
牛顿法 主要有两方面的应用：

1. 求方程的根；

2. 求解最优化方法；

一. 为什么[illegible]牛顿法求方程的根？

问题[illegible]牛顿法 是什么？目前还没有讲清楚，没关系，先直观理解为 **牛顿法是一种迭代求解方法（Newton童鞋定义的方法）**。



假设 **f(x) = 0** 为待求解方程，利用传统方法求解，牛顿法求解方程的公式：

$$f(x0+\Delta x) = f(x0) + f'(x0)\,\Delta x$$

$$\text{即 } f(x) = f(x0) + f'(x0)\,(x-x0)$$

公式可能大家会比较熟悉，一阶泰勒展式，f'(a) 表示 f(x) 在 **x0** 点的斜率 （这个很好理解），当X方向增量（**Δx**）比较小时，Y方向增量（**Δy**）可以近似表示为 斜率（导数）*X方向增量（**f'(x0) Δx**），令 f(x) = 0，我们能够得到 迭代公式：

$$x = x0 - f(x0) / f'(x0) \quad \Rightarrow \quad x_{n+1} = x_n - f(x_n) / f'(x_n)$$

通过逐次迭代，牛顿法 将逐步逼近最优值，也就是方程的解。

二. 扩展到最优化问题

这里的最优化 是指非线性最优化，解非线性最优化的方法有很多，比如 梯度下降法、共轭梯度法、变尺度法和步长加速法 等，这里我们只讲 **牛顿法。**

针对上面问题进行扩展：

解决 f(x) = 0 的问题，我们用了一阶泰勒展开：

$$f(x) = f(x0) + f'(x0)*(x-x0) + o(\,(x-x0)^2\,)$$

去掉末位高阶展开项，代入x = x0+Δx，得到：

$$f(x) = f(x0+\Delta x) = f(x0) + f'(x0)\,\Delta x$$

那么 要解决 f'(x) = 0 的问题，我们就需要二阶泰勒展开：

$$f(x) = f(x0) + f'(x0)*(x-x0) + 0.5*f''(x0)*(x-x0)^2 + o(\,(x-x0)^3\,)$$

去掉末位高阶展开项，代入x = x0+Δx，得到：

f(x) = f(x0+Δx) = f(x0) + f′(x0)Δx + 0.5 * f′′(x0) (Δx)^2

求导计算： f′(x) = f′(x0+Δx) = 0，得到：

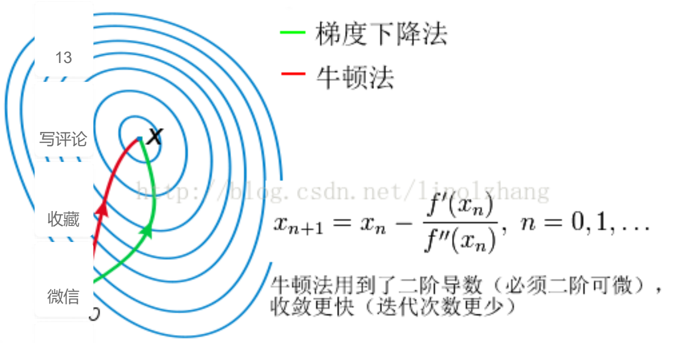
[f(x0) + f′(x0)(x−x0) + 0.5 f′′(x0)(x−x0)^2]′ = 0

整理：

f′(x0) + f′′(x0)(x−x0) = 0

x = x0 − f′(x0) / f′′(x0) => xn+1 = xn - f′(xn) / f′′(xn)

牛顿法 一图总结为：



三. 牛顿法与Hessian矩阵的关系

以上牛顿法的推导 是针对 **单变量问题**，对于多变量的情况，**牛顿法** 演变为：

$$\begin{aligned} X_{n+1} &= X_n - \frac{f'(X_n)}{f''(X_n)} = X_n - \frac{J_f(X_n)}{H(X_n)} \\ &= X_n - H^{-1}(X_n) \cdot J_f(X_n) \end{aligned}$$

与上面的单变量表示方式类似，需要用到变量的 **一阶导数 和 二阶导数。**

其中 ***J*** 定义为 **雅克比矩阵，对应一阶偏导数。**

$$J_f(X_n) = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$

H **为 Hessian矩阵，对应二阶偏导数。**

$$H(f) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1\,\partial x_n} \\ \frac{\partial^2 f}{\partial x_2\,\partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2\,\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n\,\partial x_1} & \frac{\partial^2 f}{\partial x_n\,\partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

网上也能搜到类似的公式表达，也列出来：

$$x_{n+1} = x_n - [Hf(x_n)]^{-1}\nabla f(x_n), n \geq 0$$

牛顿法 在多变量问题上仍然适用迭代求解，但Hessian矩阵的引入增加了复杂性，特别是当：

- **Hessian 矩阵非正定（非凸）导致无法收敛；**
- **Hessian 矩阵维度过大带来巨大的计算量。**

针对这个问题，在 牛顿法无法有效执行的情况下，提出了很多改进方法，比如 **拟牛顿法**（Quasi-Newton Methods）可以看作是牛顿法的近似。

拟牛顿法 只需要用到一阶导数，不需要计算Hessian矩阵 以及逆矩阵，因此能够更快收敛，关于 **拟牛顿法** 这里不再具体展开，也有更深入的DFP、BFGS、L-BFGS等算法，大家可以自行搜索学习。

总体来讲，拟牛顿法 都是用来解决 牛顿法 本身的 复杂计算、难以收敛、局部最小值等问题。

文章标签： 牛顿法 Hessian矩阵 雅可比矩阵 最小二乘法 非线性优化

个人分类： 机器学习 | 计算机视觉 数学

相关热词： 牛顿法你牛顿法 牛顿法高斯牛顿法 牛顿法阻尼牛顿法 一阶牛顿法 二分法和牛顿法

上一篇 涉 习之对抗网络 下一篇 奇异值分解SVD

想对 点什么

昊86 05-22 22:41:00 #2楼

问个问 如下原文： 去掉末位高阶展开项，代入x = x0+Δx，得到： f(x) = f(x0+Δx) = f(x0) + f′(x0)Δx + 0.5 * f″(x0) (Δx)^2 求导计算： f′(x) = f′(x0+Δx) = 0，得到： f′(x0) + f′(x0)(x−x0) + 0.5 f″(x0)(x−x0)^2]′ = 0 -----这是怎么得到的呀？ 看起来是让f(x)=0变换得到的呀。 整理： f′(x0) + f″(x0)(x−x0) = 0 这个又是怎样整理得到的？ 谢谢啊。

Velvas_ 2018-03-07 19:55:45 #1楼

受益匪浅，感谢

上一页 1 下一页

牛顿法以及雅克比矩阵、海森矩阵（Hessian）数学方法。

1531

一般来说, 牛顿法主要应用在两个方面, 1, 求方程的根; 2, 最优化。 1，求方程的根 其原理便是使用泰勒展开，然后去线性部分…

Hessian的基本使用

1191

完整代码下载 https://gitee.com/zml2015/HessianDemo 引入必要jar包 com.caucho hessian 4.0.51 ...

hessian矩阵特征值 - CSDN博客

5-25

求hession矩阵的特征值和特征向量 //求hession矩阵的特征值和特征向量 void CmCurveEx::compute_eigenvals(double dfdrr, double df…

Hessian 矩阵(黑塞矩阵)以及hessian矩阵奇异的用法 - CSDN博客

7-18

求解方程f(x)=0,即f(x0)+(x-x0)*f′(x0)=0,求解x = x1=x0-f(x0)/f′(x0),因为这是利用泰勒公式的一阶展开,f(x) = f(x0)+(x-x0)...

Hessian 矩阵（黑塞矩阵）以及hessian矩阵奇异的用法

1.3万

Hessian Matrix（黑塞矩阵、海森矩阵、海瑟矩阵、海塞矩阵等）

海森矩阵 Hessian matrix - CSDN博客

7-2

求hession矩阵的特征值和特征向量 //求hession矩阵的特征值和特征向量 void CmCurveEx::compute_eigenvals(double dfdrr, double df…