

为什么梯度是函数变化最快的方向

原创 石头 机器学习算法那些事 2018-11-05

前言

机器学习中的大部分问题都是优化问题，而绝大部分优化问题都可以用梯度下降法来解决。本文详细的解释了高数中几个易混淆的重要概念，如导数和微分的区别，偏导数的概念，方向导数和梯度的关系，若完全掌握这几个概念，就能很好的理解梯度为什么是函数变化最快的方向的问题。

目录

- 1、导数和微分
- 2、偏导数
- 3、方向导数和梯度的关系
- 4、总结

导数和微分

导数的定义

定义：设函数 $y = f(x)$ 在 x_0 领域内有定义，当自变量 x 在 x_0 处有增益 Δy ，若 $\frac{\Delta y}{\Delta x}$ 在 $\Delta x \rightarrow 0$ 极限存在，则称函数 $f(x)$ 在该点可导，记为 $f'(x_0)$ ，表达式如下：

$$f'(x_0) = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

本质：导数描述的是函数在一点处的变化快慢的趋势，是一个变化的速率。如曲线方程的导数是随点变化的斜率，运动方程的导数是随时间变化的速率。

微分的定义

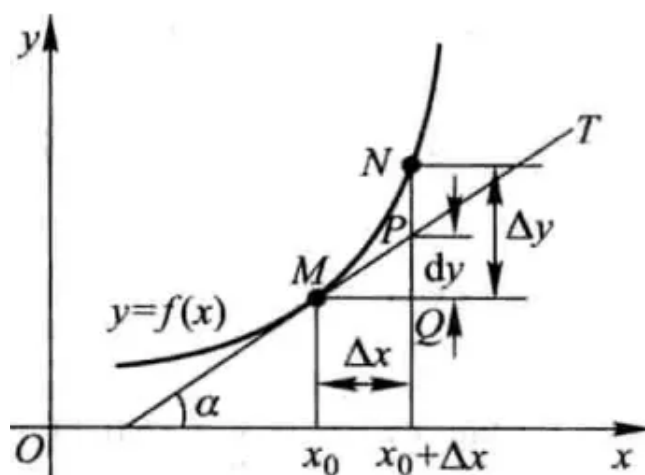
定义：函数 $y = f(x)$ 在 x_0 领域内 $[x_0, x_0 + \Delta x]$ 有定义，对应的函数增量 $\Delta y = f(x_0 + \Delta x) - f(x_0)$ 。若函数增量可表示为： $\Delta y = A\Delta x + o(\Delta x)$ ，其中 A 是不依赖于 Δx 的常数， $o(\Delta x)$ 是 Δx 的高阶无穷小则称函数是可微的，其中 $A\Delta x$ 称为微分，记为 dy

本质：微分描述的是函数从一个点移动到另一个无穷小的点所产生的的变化量。

函数增量与微分的关系

本节分别从图形角度和代数角度去分析函数增量与微分的关系：

○ 图形角度：



如上图所示，函数 $f(x)$ 在 M 点处的导数为直线 T 的斜率 $\tan \alpha$ ， Δy 是 M 点移动 Δx 时的函数增量， dy 为函数相对于 Δx 的微分。

当 $\Delta x \rightarrow 0$ 时， $dy \approx \Delta y$ ， $dy = (\tan \alpha) * \Delta x$ 。

○ 代数角度：

若 $f(x)$ 满足微分条件，则

$\Delta y = A\Delta x + o(\Delta x)$ ，两边同时除以 Δx ，得：

$$\frac{\Delta y}{\Delta x} = A + o(\Delta x)$$

当 $\Delta x \rightarrow 0$ 时，

$$A = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = f'(x_0)$$

当 $\Delta x \rightarrow 0$ 时， $dy \cong \Delta y$

偏导数

偏导数是函数相对于某一轴方向的导数，其他轴方向则假设为常数，若考虑二元变量 $f(x,y)$ ，偏导数定义如下：

如果：

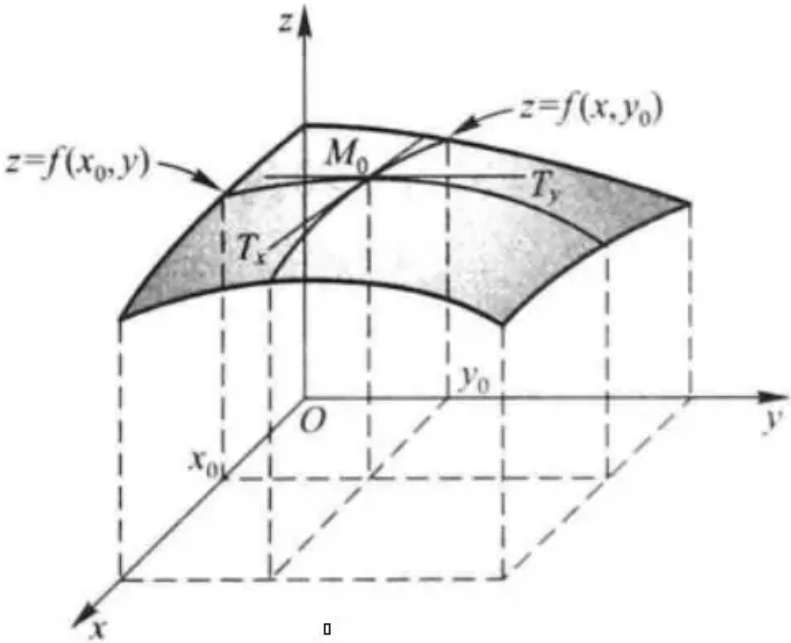
$$\lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x, y_0) - f(x_0, y_0)}{\Delta x}$$

存在，则称该式为 $f(x,y)$ 在 (x_0, y_0) 对 x 的偏导数， $f(x,y)$ 在 (x_0, y_0) 对 y 的偏导数可用同样的方法求解。

偏导数的几何意义

令 $z = f(x, y)$ ，偏导数 $f_x(x_0, y_0)$ 等价于曲面被片面 $y = y_0$ 所截得的曲线在点 M_0 处的切线 M_0T_x 对 X 轴的斜率，同理偏导数 $f_y(x_0, y_0)$ 等价于曲面被平面 $x = x_0$ 所截得的曲线在点 M_0 处的切线 M_0T_y 对 Y 轴的斜率

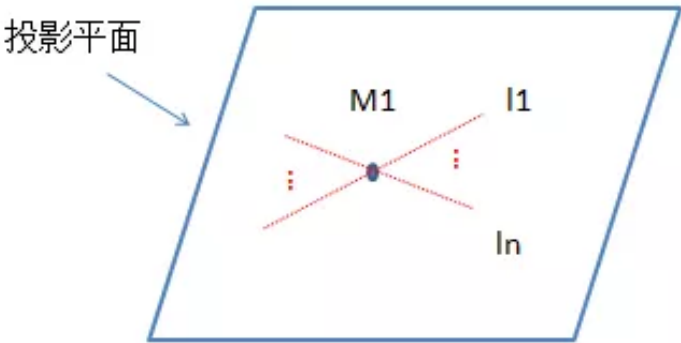
如下图：



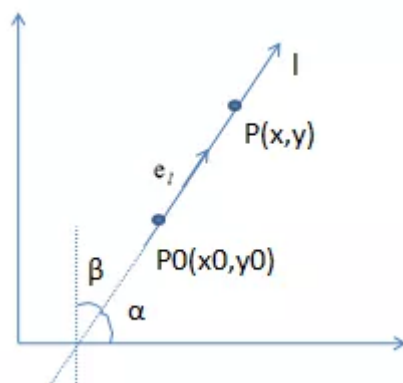
方向导数和梯度的关系

方向导数

我们还是以讨论偏导数的图来解释方向导数。令曲面方程 $z=f(x,y)$ 投影到XY平面，得到投影平面，如下图：



M1为M0在XY面的投影点，由上图可知，有无数条直线经过M1点，这些直线代表方向，我们认为曲面M1点的方向导数就是求这些直线方向的导数，M1点的方向导数也是无穷多个，我们用变量 α 来代表不同的方向直线。



如上图，直线 l 的方向向量 $\vec{e}_l = (\cos \alpha, \cos \beta)$ ， t 为 $|P_0P|$ 的距离
直线 l 的参数方程：

$$\begin{cases} x = x_0 + t \cos \alpha \\ y = y_0 + t \cos \beta \end{cases}$$

$\therefore f(x, y)$ 在 M_0 点沿方向向量 \vec{e}_l 的方向导数为：

$$\frac{\partial f}{\partial l} \Big|_{(x_0, y_0)} = \lim_{t \rightarrow 0^+} \frac{f(x_0 + t \cos \alpha, y_0 + t \cos \beta) - f(x_0, y_0)}{t}$$

由上式可知，方向导数随夹角 α 不同而不同

由第一节介绍的单元变量的微分公式可推导二元变量的全微分公式

$$f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0) = f_x(x_0, y_0)\Delta x + f_y(x_0, y_0)\Delta y + o(t) \quad (1)$$

其中， $t = \sqrt{(\Delta x)^2 + (\Delta y)^2}$ ， $t \cos \alpha = \Delta x$ ， $t \cos \beta = \Delta y$

当 $t \rightarrow 0$ 时，(1)式两边各除 t ，得：

$$\lim_{t \rightarrow 0} \frac{f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0)}{t} = f_x(x_0, y_0) \cos \alpha + f_y(x_0, y_0) \cos \beta$$

由方向导数的定义可知：

$$\frac{\partial f}{\partial l} \Big|_{(x_0, y_0)} = f_x(x_0, y_0) \cos \alpha + f_y(x_0, y_0) \cos \beta$$

梯度

梯度是一个矢量，曲面上每点的梯度是常数， P_0 点的梯度如下：

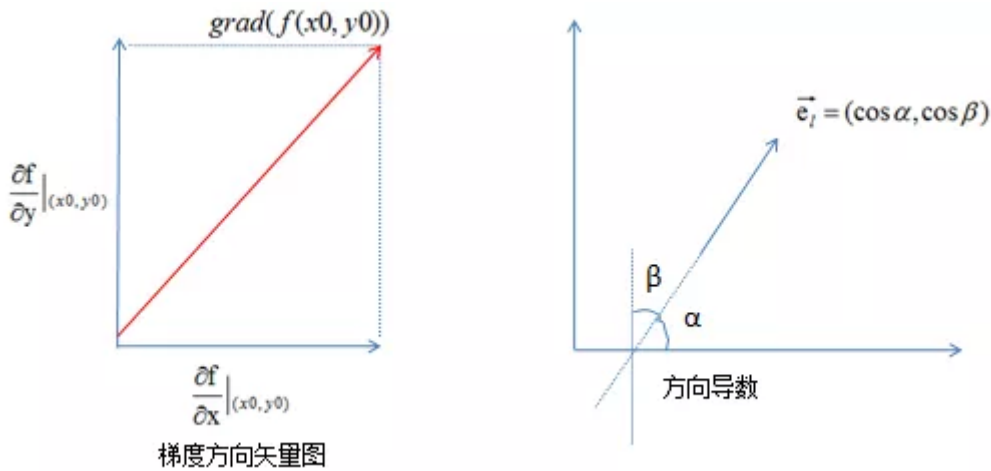
$$\text{grad}(f(x_0, y_0)) = f_x(x_0, y_0) * \vec{i} + f_y(x_0, y_0) * \vec{j}$$

其中 \vec{i}, \vec{j} 分别表示 x 轴和 y 轴的方向

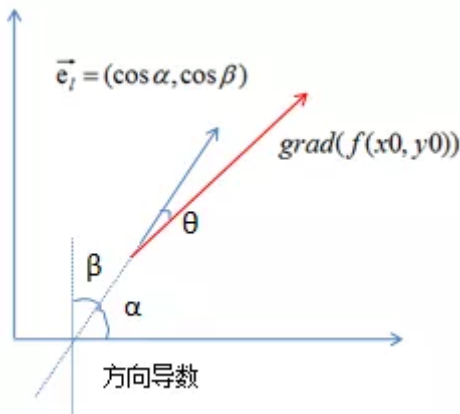
方向导数和梯度的关系

求上图曲面M0中P点的梯度和方向导数

梯度和方向导数的单位向量分别如下两图：



平移梯度向量，使之与方向导数的单位向量相交，夹角为 θ ，如下图：



红色直线代表梯度，蓝色代表方向导数的单位向量，取该两个向量的内积，得：

$$\text{grad}(f(x_0, y_0)) \cdot \vec{e}_l = (f_x(x_0, y_0), f_y(x_0, y_0)) \cdot (\cos \alpha, \cos \beta)$$

$$\text{grad}(f(x_0, y_0)) \cdot \vec{e}_l = f_x(x_0, y_0) \cos \alpha + f_y(x_0, y_0) \cos \beta$$

由方向导数的表达式可知：

$$\left. \frac{\partial f}{\partial l} \right|_{(x_0, y_0)} = \text{grad}(f(x_0, y_0)) \cdot \vec{e}_l$$

$$\left. \frac{\partial f}{\partial l} \right|_{(x_0, y_0)} = |\text{grad}(f(x_0, y_0))| \cdot \cos \theta$$

$\therefore \theta = 0$ 时，方向导数等于梯度，且取得最大值

结论：曲面中点的方向导数有无数个，当方向导数与梯度方向一致时，该导数值取得最大，等价于该点在梯度方向具有最快的变化值。梯度方向是函数值增加最快的方向，梯度的反方向是函数值减小最快的方向。

总结

本文介绍了高数教材中几个易混淆的概念，结合图解法和公式推导法证明了方向导数和梯度方向一致时，函数值是变化最快的。因此，机器学习常用梯度法去解决最优化问题。

参考

《高等数学》第七版，同济大学

推荐阅读文章

线性分类模型（二）：logistic回归模型分析

线性分类模型（一）：浅判别模型分析

深入理解线性回归算法（三）：浅谈贝叶斯线性回归

深入理解线性回归算法（二）：正则项的详细分析

深入理解线性回归算法（一）



-END-



长按二维码关注

机器学习算法那些事
微信: beautifulife244

砥砺前行 不忘初心