函数的切线与梯度

Dezeming Family

2021 年 8 月 30 日

DezemingFamily 系列书和小册子因为是电子书，所以可以很方便地进行修改和重新发布。如果您获得了 DezemingFamily 的系列书，可以从我们的网站 [https://dezeming.top/] 找到最新版。对书的内容建议和出现的错误欢迎在网站留言。

20210831：完成第一版。

20210902：第二版增加了关于梯度和法向量关系的内容。

20210915：第三版增加了关于梯度是函数值上升的方向的相关原理描述。

目录

一 一元函数的切线与梯度

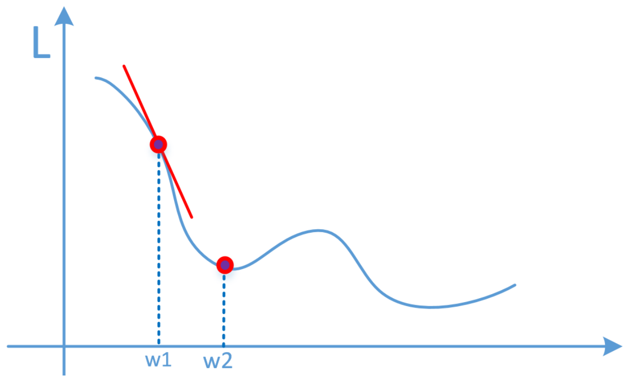
当考虑到一元函数时，因为可以在平面图上表示，所以简单一些。

我们知道，一元函数的导数就是当前点切线的斜率，即切线 $y = kx + b$ 中的 k。

至于梯度，我们先举个例子。在机器学习中，有一个术语叫做梯度下降法，其实就是找函数的极小值点。假如我们有 10 个样本，得到的线性回归损失函数就是（大家不需要知道什么是损失函数和机器学习，只需要注意到我们的公式形式）：

$$L(w,b) = \sum_{i=1}^{10} \Big(y_i - (b + w \cdot x_i)\Big)^2 \tag{一.1}$$

我们的目标是变化 w，使得 L 函数的值尽可能小。对于不同的 w，我们可以绘制一个 L 函数曲线：

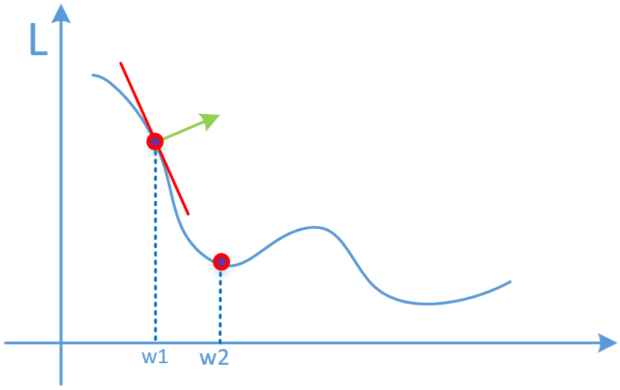


虽然 w_2 并不是整个函数的最小值处，但至少比 w_1 处的值要小，因此我们需要想办法让 w 从 w_1 处移动到 w_2 处。

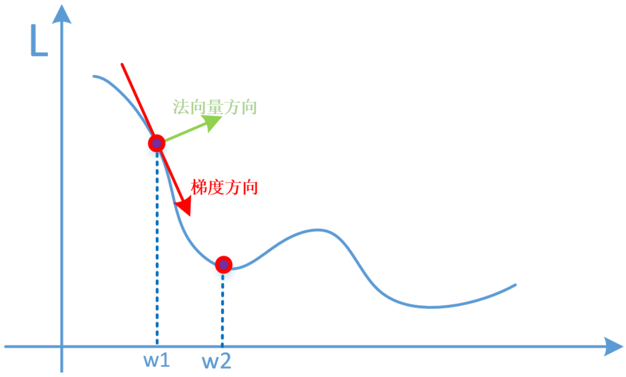
根据梯度下降方法的原理，我们可以得到向函数值更小的方向移动的公式：$w_1' \leftarrow w_1 - \eta \frac{dL}{dw}\Big|_{w=w_1}$。我们知道的是，$\frac{dL}{dw}$ 是函数在当前点的斜率，当 k 大于 0 的时候，说明函数在当前点随着 w 变大是上升趋势，所以应该向反方向走，反之亦然。

因此我们知道：单变量函数里，函数在某个点的梯度就是函数在当前点切线的斜率。

现在再思考一下法向量，即下图的绿色向量：



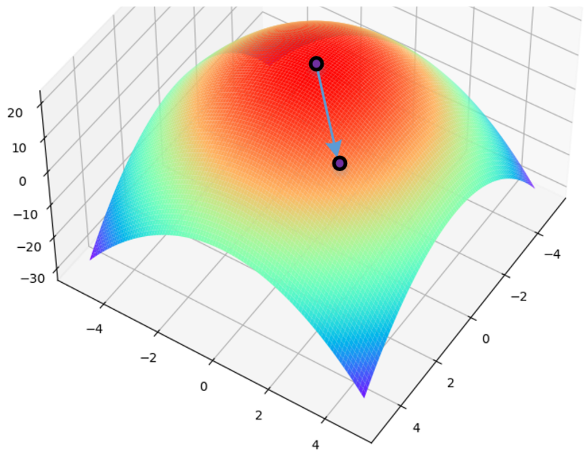
当切线不考虑截距时，表示为 $y = kx$，即方向向量为 (n, kn)（切线是没有方向的，我们这里表示的方向向量是与切线平行的向量）。也就是说法向量相应地为 $(kn, -n)$（要与切线方向向量垂直，至于法向量朝里还是朝外，是与 n 有关，前面的方向向量的方向也与 n 有关）。一维函数的梯度就是导数，它代表了函数增加的方向。当我们把 $(x, f(x))$ 当做一条曲线时，就能根据切线来计算出法向量（当然，具体情况还是要以不同的资料定义的内容为准，但基本原理都是一样的）：



接下来我们探讨一下多元函数的切线与梯度。

二　多元函数的切线与梯度

对于多元函数来说，例如二元函数，函数在某个点上的切线不只有一条，所有切线构成一个切面。多元函数的梯度下降类似于一元函数，同样是移动参数使得 L 函数值尽可能地小：



2 1 切平面的求法

设二元函数 $z = f(x, y)$，在某点 (x_0, y_0) 处的切平面可以表示为：

$$[f'_x(x_0, y_0)](x - x_0) + [f'_y(x_0, y_0)](y - y_0) = (z - z_0) \tag{二.1}$$

其中，f'_x 和 f'_y 分别是函数在 x 和 y 方向的偏导。

证明：

首先我们研究的内容是，已知一个平面上的点和平面法向量，求该平面的公式。假如我们有平面上一个点 (x_0, y_0, z_0)，以及该平面的法向量 (x', y', z')，因为法向量垂直于平面，所以设平面另外一个点为 (x, y, z)，于是 $(x - x_0, y - y_0, z - z_0)$ 将与 (x', y', z') 垂直，即：

$$(x - x_0) \cdot x' + (y - y_0) \cdot y' + (z - z_0) \cdot z' = 0 \tag{二.2}$$

设函数 $\psi = z - f(x, y) = 0$，该函数在某点的切面法向量方向向量为：

$$\left(\psi'_x, \psi'_y, \psi'_z\right) \tag{二.3}$$

可以求出：

$$\begin{cases} \psi'_x = -f'_x(x_0, y_0) \\ \psi'_y = -f'_y(x_0, y_0) \\ \psi'_z = 1 \end{cases} \tag{二.4}$$

因此切面的方程就是：

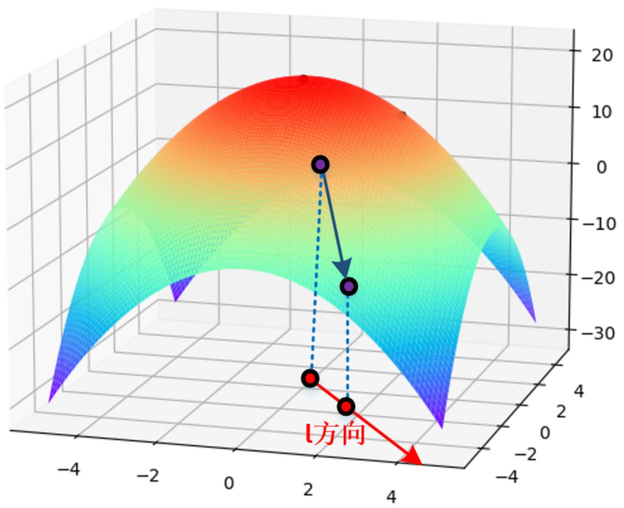
$$-f'_x(x_0, y_0)(x - x_0) - f'_y(x_0, y_0)(y - y_0) + (z - z_0) = 0 \tag{二.5}$$

2 2 函数的方向导数

对于函数 $z = z(x, y)$，设有一个点 $p_0 = (x_0, y_0)$，设现在有另一点 p，沿着从 p_0 引出的某条射线 l 上移动，当 p 逐渐靠近 p_0 时，若下面的极限存在，则称为函数在该点 p_0 处的方向导数。

$$\left.\frac{\partial z}{\partial l}\right|_{p_0} = \lim_{p \to p_0} \frac{z(p) - z(p_0)}{\overline{p_0 p}} \tag{二.6}$$

其中 $\overline{p_0 p}$ 代表的是两点间的距离。可以看出，当 $\frac{\partial z}{\partial l} > 0$ 表示函数值沿着 l 方向增加；$\frac{\partial z}{\partial l} < 0$ 表示函数值沿着 l 方向降低。如下图的红线，即表示函数变化方向 l。



现在我们虽然知道怎么求方向导数，但我们想知道的是方向导数与偏导 $\frac{\partial z}{\partial x}$、$\frac{\partial z}{\partial y}$ 之间的关系。

设动点 p 的坐标为 $p(x_0 + \Delta x, y_0 + \Delta y)$，因此（见《函数连续、可微、可导之间的关系》）：

$$\Delta z = z(p) - z(p_0) \tag{二.7}$$

$$= \frac{\partial z}{\partial x}\Delta x + \frac{\partial z}{\partial y}\Delta y + o(\overline{p_0 p}) \tag{二.8}$$

其中 $o(\overline{p_0p})$ 是 $\overline{p_0p}$ 的高阶无穷小。等式两边同时除以 $\overline{p_0p}$，得到：

$$\frac{\Delta z}{\overline{p_0p}} = \frac{\partial z}{\partial x}\frac{\Delta x}{\overline{p_0p}} + \frac{\partial z}{\partial y}\frac{\Delta y}{\overline{p_0p}} + \frac{o(\overline{p_0p})}{\overline{p_0p}} \tag{二.9}$$

所以当 $\overline{p_0p} \to 0$ 时就可以得到：

$$\frac{\partial z}{\partial l} = \frac{\partial z}{\partial x}cos\alpha + \frac{\partial z}{\partial y}cos\beta \tag{二.10}$$

其中 α 和 β 为方向向量。

2 3 函数梯度

二元函数 $z = f(x,y)$ 在某点 (x_0,y_0) 的梯度表示为 $\nabla z = \Big(f'_x(x_0,y_0), f'_y(x_0,y_0)\Big)$。

梯度表示方向是函数值变化率最大的方向，证明如下。

我们设 l 方向的单位矢量为 $\boldsymbol{l} = cos\alpha\boldsymbol{i} + cos\beta\boldsymbol{j}$，方向导数的偏导项也可以视为一个矢量 $\boldsymbol{G} = \frac{\partial z}{\partial x}\boldsymbol{i} + \frac{\partial z}{\partial y}\boldsymbol{j}$，也就是说方向导数可以写作：$\frac{z}{l} = \boldsymbol{G}\cdot\boldsymbol{l}$，即：

$$\frac{\partial z}{\partial l} = \big|\boldsymbol{G}\big|cos\big(\boldsymbol{G},\boldsymbol{l}\big) \tag{二.11}$$

也就说明，当 $\boldsymbol{l}$ 和 $\boldsymbol{G}$ 在同一个方向时，方向导数可以取到最大值。因此矢量 $(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y})$ 就是函数变化率最大的方向。我们称这个方向为梯度方向。

我们来分析一下梯度方向是函数上升最快的方向还是下降最快的方向。

一元函数梯度很明显是函数值上式最快的方向，在第一节就有描述。

对于多元函数，比如二元函数 $z = f(x,y)$，我们在某个点 (x_0,y_0) 处固定某个变量，例如固定 $y = y_0$，然后单看 x 维度上的函数 $z = f(x,y_0) = g(x)$，在 $y = y_0$ 时的偏导 $f'(x,y_0) = g'(x)$，偏导代表了沿着 x 方向函数上升的方向，固定 $x = x_0$ 也是同理。所以说，**多元函数的梯度代表了函数值上升最快的方向**。

2 4 区分梯度与法向量

我们要注意，曲面的法向量表示和多元函数梯度是一致的，这句话我们需要这么理解：只有函数才会有梯度，表示函数的最大变化方向；曲面没有梯度，但是曲面有法向量。

对于一元函数 $y = f(x)$，梯度由求导得到。即 $f'(x)$，

对于二元函数 $z = f(x,y)$，梯度是变化最快（函数上升最快）的方向，计算为 $(f'_x(x,y), f'_y(x,y))$。

对于三元函数 $s = f(x,y,z)$，梯度是 (f'_x, f'_y, f'_z)。

而曲面其实相当于比函数的定义低了一维，例如二元函数转为曲面，就是 $f(x,y) - z = 0$，我们可以设曲面表达式为 $\Phi(x,y,z) = f(x,y) - z = 0$，计算得到曲面法向量为 $(\Phi'_x, \Phi'_y, \Phi'_z)$。其实这个法向量就是 $(f'_x, f'_y, -1)$。

参考文献

[1] 谢树艺. 工程数学矢量分析与场论 [M]. 高等教育出版社, 2012.

[2] https://www.zhihu.com/question/63792598/answer/654061538