

多参数的链式法则

Dezeming Family

2022 年 2 月 2 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。

目录

一 介绍	1
二 多参数	1
三 证明	3
参考文献	3

一 介绍

本文来自于我在学习 miniTorch[1] 项目的自动微分时遇到的问题，多参数也是 [1] 中的描述方法（实际为多元复合导数）。

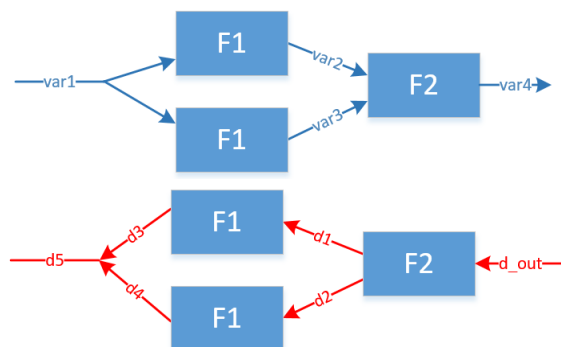
虽然这是微积分中必学的知识点，但是本人由于当初本科课程并没有涉及到这些内容，所以甚至没有任何的印象，并不清楚这些知识，而是自己花了不少时间推导，感叹自己知识储备不足的同时，在这里进行一下记录和描述来引以为戒。

本文的介绍依赖于 [1] 的过程和 [2] 中关于多元复合函数导数的理论。

二 多参数

例子 1

我们先把计算图画出来（上面表示 forward() 过程，下面表示 backward() 过程）：



其中, var 表示变量, F 表示函数。我们设:

$$F1(x) = 2x \quad (二.1)$$

$$F2(x, y) = x * y \quad (二.2)$$

因此:

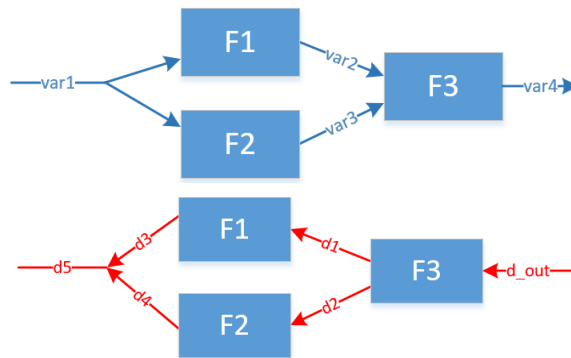
$$var4 = F2(var2, var3) = F2(F1(var1), F1(var1)) \quad (二.3)$$

先考虑正向过程, 设 $var1 = 2$, 则 $var2 = var3 = 4$, $var4 = var2 * var3 = 16$ 。

设反向传播的 $d_out = 1$, 则 $d1 = \frac{d(var2*var3)}{d var2} = var3 = 4$, 同理 $d2 = 4$; $d3 = 4 * \frac{d 2*var2}{d var2} = 8$, $d4 = 8$ 。现在的关键是变量 $var1$ 的导数, 也就是 $d5$ 是多少? 可能一开始有人会觉得 (比如我) $d5 = d3 = d4 = 8$ 。我们先看下面的这个例子。

例子 2

我们再换个例子:



我们设:

$$F1(x) = 2x \quad (二.4)$$

$$F2(x) = 3x + 1 \quad (二.5)$$

$$F2(x, y) = x * y \quad (二.6)$$

先考虑正向过程, 设 $var1 = 2$, 则 $var2 = 4$, $var3 = 7$, $var4 = var2 * var3 = 28$ 。

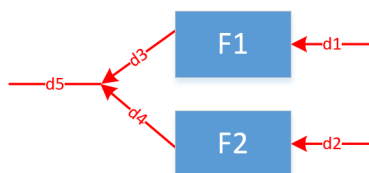
设反向传播的 $d_out = 1$, 则 $d1 = \frac{d(var2*var3)}{d var2} = var3 = 7$, $d2 = 4$; $d3 = 7 * \frac{d 2*var2}{d var2} = 14$, $d4 = 12$ 。此时 $d3 \neq d4$, 因此, $d5 = d4 = d3$ 是肯定不对的。

实际结果

针对例子 1, $var4 = 2(var1) * 2(var1) = 4(var1)^2$, 因此, $\frac{d var4}{d var1} = 8var1 = 16$ 。 $16 = 8 + 8$ 。

针对例子 2, $var4 = 2(var1) * (3(var1) + 1) = 6(var1)^2 + 2(var1)$, 因此, $\frac{d var4}{d var1} = 12var1 + 2 = 26$ 。
 $26 = 14 + 12$ 。

我们可以看到, 分叉处的导数就是两边的导数的和。这也很好理解, 我们可以想象, 设函数一开始时 $F1(x) \equiv F2(x) = 2x$; 然后当 $F2$ 变为 $F2(x) = 3x$ 后, 整个函数的求导对于 $var1$ 来说肯定会变大。在这里, $var1$ 相当于分开了两条支路, 各自去进行计算:



三 证明

对于一元函数链式法则，可以描述如下：

$$\frac{d F2(F1(x))}{d x} = \frac{d F1(x)}{d x} \times \frac{d F2(F1(x))}{d F1(x)} \quad (\text{三.1})$$

我们当前的情景为多元复合函数，求其微积分的方式可见 [2]：

$$\frac{d F3(F1(x), F2(x))}{d x} = \frac{d F3(F1(x), F2(x))}{d F1(x)} \times \frac{d F1(x)}{d x} + \frac{d F3(F1(x), F2(x))}{d F2(x)} \times \frac{d F2(x)}{d x} \quad (\text{三.2})$$

参考文献

[1] <https://minitorch.github.io/backpropagate.html>

[2] 《高等数学第 7 版》下册同济大学