

微积分

目录

- 概述
 - 知识点
 - 微分(Derivative)
 - 偏微分(Partial Derivative)
 - 梯度(Gradient)
 - 练习
 - 拓展资源(Optional)
 - 链式法则(Chain Rule)
 - 常用微积分公式
-

1. 概述

在这一部分，你将接触到机器学习&人工智能领域所需要的微积分知识。有一些内容对于你来说可能已经相当熟悉，我们希望能唤起你（希望不是很久远）的记忆。作为传统机器学习优化方法的基石，微积分将是我们课程学习最重要的基础之一。

2. 微分(Derivative)

- 微分的本质 & 概念
 - 视频讲解（只需要看 Part1&Part2）：
<https://www.bilibili.com/video/BV1qW411N7FU?p=1>
 - 微分本质上是反映函数在自变量（X）极小变化下，因变量（Y）是如何改变的。
- 微分公式 & 应用
 - 参考视频：<https://www.bilibili.com/video/BV1qW411N7FU?p=3>

$f(x) = x^n$	$\frac{df}{dx} = nx^{n-1}$	想了解公式推导，可观看参考视频P3
$f(x) = \sin x$	$\frac{df}{dx} = \cos x$	参考视频P3 12:40
$f(x) = \cos x$	$\frac{df}{dx} = -\sin x$	
$f(x) = a^x$	$\frac{df}{dx} = a^x \ln a$	参考视频P5
$f(x) = e^x$	$\frac{df}{dx} = e^x$	参考视频P5
$f(x) = \ln x$	$\frac{df}{dx} = \frac{1}{x}$	

- 二阶导数

- 对一个函数的导数求导便是二阶导数

- 举个栗子: $y = 3x^2 - 2e^{2x} + 2\sin(x)$

- $\frac{dy}{dx} = 6x - 4e^{2x} + 2\cos(x)$

- $\frac{d^2y}{dx^2} = 6 - 8e^{2x} - 2\sin(x) \leq$ 这就是原函数 y 的二阶导数!

3. 偏微分(Partial Derivative)

- 多元微积分(Multivariable Calculus)

- 在很多情况下, 我们关心的因变量 (Y) 可能由多个变量同时影响。在此时, 我们可能会有如下的表达式: $y = 3x_1 - 2e^{x_2} + \sin^2(x_3)$

- 在这一情况下, x_1, x_2, x_3 发生微小变化时 y 的变化量自然是各有不同, 那我们如何描述这一变化量呢? 这就得引入多元微积分 —— 也就是有多于一个自变量 (X) 时的微积分操作。在这一环节, 我们只会关心其中一个概念: 偏微分 (Partial Derivative)

- 偏微分的概念

- 视频讲解: <https://www.bilibili.com/video/BV1L7411a7Xk>

- 偏微分的计算有两个核心元素: 函数表达式 ($y = \dots$) 和变量 (其中一个变量)

- 举个栗子: $y = 3x_1x_2 - 2e^{x_2} + \sin^2(x_3)$

- 在计算 y 对于 x_2 的偏微分时, 先将 x_1, x_3 暂时看作“常数” (虽然他们并不是)。注意: 是 y 对于 x_2

- 然后就可以像平时求微分一样计算了!

- 结果是: $\frac{\partial y}{\partial x_2} = 3x_1 - 2e^{x_2}$

4. 梯度(Gradient)

- 不知道大家有没有听说过“梯度下降”这个词? 作为机器学习领域最重要的一种优化 (Optimization) 方法, 这一方法广泛应用于大量机器学习的模型中。而这一方法的理论基础就是“梯度”(Gradient)
- 视频讲解: <https://www.bilibili.com/video/BV11a4y1L7BV?p=19> (只需看Part19前3分15秒, 以及Part20)
- Gradient 其实并不是一个新的概念, 我们可以来看一下它的数学表达式:

- $gradf = (\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n})$ ($1 * n$ vector)
- 然后让我们来看一眼“正常”的微分
 - $Df = [\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}]$ ($n * 1$ matrix)
- 两者的内容没有任何区别！只是“形状”稍有不同(梯度是一个 $n * 1$ 的向量而多元微分是一个 $1 * n$ 的“长条”矩阵)，我们暂时不需要知道这一形状差异为他们带来的差别。
- 所以，长话短说：梯度就是把多元函数按着自变量顺序一一求偏导，然后“叠”起来组成的一个向量！

5. 练习

1. Q1(微分)对如下函数求二阶导数：

$$f(x) = e^{2x} + \sin(x^2)$$

2. Q2(偏微分)对如下函数求对 x_1 的偏导：

$$f(x_1, x_2) = 2e^{2x_1} + 3x_1x_2 + \sin(x_2^2)$$

3. Q3(梯度)请写出如下函数的梯度：

$$f(x_1, x_2) = 2e^{2x_1} + 3x_1x_2 - \ln(x_2)$$

6. 拓展阅读 (Optional)

- 链式法则 (Chain Rule)
 - 视频讲解 (Part4) <https://www.bilibili.com/video/BV1qW411N7FU?p=4>
- 常用微积分公式法则
 - <https://techx.feishu.cn/file/boxcndezp9WygQ28aVwbFkmz4eg>