

优化 / Optimization

目录

- 概述
- 知识点
 - **Loss Function**
 - **Gradient**
- 练习
- 拓展资源 (Optional)
 - 牛顿法
 - 一篇不错的论文

1. 概述

- 今天预习的内容是优化，优化的意思是“我们如何得到更好（甚至最好）的模型”。在这一部分中，我们将先学习如何定义“好”的模型 —— Loss & Metrics；然后我们将重新回顾一个微积分方面的概念 —— 梯度，以及大名鼎鼎的优化发：“梯度下降法”。

2. Loss Function

- 在介绍概念之前，我们先从比较广义的角度来看一下机器学习问题：在机器学习中，我们往往假设“我们可以用一种有条理的方式（模型）系统地认识或预测数据在这个过程中，模型的选择和模型内的参数都是未知的，这也意味着我们需要从茫茫“模型海”中找到最贴近现实数据的那个模型 —— 也就是说，我们希望的“好模型”应该是“最贴近现实数据的模型”。接下来我们将用 “Loss Function” 这个概念来定义什么是“贴近现实数据”。
- **Loss Function**, 又称“损失函数”，是机器学习训练过程中的一个重要概念。它意味着训练的时候，我们的模型与“完美模型”之间的差距。举个栗子：
 - 我们有 100 个同学的【语文，英语】成绩（假设分别为 $x_1, \dots, x_{100}, y_1, \dots, y_{100}$ ）。此时，我们希望能找到他们语文英语成绩之间的关系，我们选择了线性模型： $\hat{y} = a_1 + a_2 * x$ 。我们训练的过程也就是寻找一对 (a_1, a_2) 使得这个模型预测的 \hat{y} 与真实的 y 差距尽可能小。
 - 那么问题来了，这个“差距”应该如何表示呢？

- 在这个栗子中，我们会用 MSE (Mean Square Error) 来表达 

$$MSE = \frac{1}{100} * \sum_{i=1}^{100} (y - \hat{y})^2$$

- 可以从公示中看出，MSE 的数值可以比较精确地衡量我们的模型在整体上是否能准确预测，假如对于某些同学预测得很好、但对另外的同学预测得不好，它的数值仍然会很大。

3. 梯度 (Gradient)

- 回顾一下我们在“微积分”一章中提到的“梯度”那一概念：

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)

- 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：

$$\circ \quad \text{grad} f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right) \quad (1 * n \text{ vector})$$

- 然后让我们来看一眼“正常”的微分

$$\circ \quad Df = \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right] \quad (n * 1 \text{ matrix})$$

- 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！

- 当时我们没有介绍梯度的应用价值：“一个在 n 元函数上的点 P 对应的梯度 (x_1', x_2', \dots, x_n')，作为一个向量、它指向方向是这个函数在最大增长的方向、而它的量(大小)是在这个方向上的增长率 (摘自维基百科)”。对于这个特点，我们暂时不花时间进行数学证明，请大家先接受这一关于梯度的特点。
- 那么梯度下降又是什么呢？我们找到了一个非常形象的视频来解释 (绝不是在偷懒)：参考视频<https://www.bilibili.com/video/BV1eC4y187yf>

4. 拓展阅读

- 常见的损失函数 (Loss Function) :
<https://medium.com/data-science-group-iitr/loss-functions-and-optimization-algorithms-demystified-bb92daff331c>
- 牛顿法
 - 视频讲解：
<https://www.bilibili.com/video/BV1Zk4y1q7qP?from=search&seid=7571157521557084884>

- [一篇不错的论文](#)