

- 梯度下降法
 - 目的
 - 方法

梯度下降法

目的

找到使得成本函数 $J(w, b)$ 最小的参数 w 和 b

方法

首先我们应当选择一个值对 w, b 进行初始化

我们很容易看到,使用上一章所说的成本函数进行计算的成本函数 $J(w, b)$ 是一个凸函数,所以无论选择那个一点进行初始化实际上都是一样的,最终都会到达某一个特定的点,或者十分接近这个点的附近 通常选取0作为初始值

过程:从初始点开始,最陡的下坡方向(也就是梯度最大的一个方向走)走一步,称为一次迭代,在若干次迭代后, w, b 的值会到达或者无限接近于一个全局最优解

我们可以先从二维视角来观察,取 wOJ 面,假设一次迭代的结果为

$$w' = w - \alpha \frac{dJ(w)}{dw} \quad (\alpha \text{为学习率,可以控制每一次迭代的步长})$$

在算法收敛前,不断重复地进行这一个过程,最终达到或无限接近全局最优解

但实际上成本函数含有 w, b 两个参数,所以需要不断地通过

$$w' = w - \alpha \frac{dJ(w, b)}{dw}$$
$$b' = b - \alpha \frac{dJ(w, b)}{db}$$

这个过程来进行更新