梯度下降法学习

# 了解什么是梯度

参考高等数学第七节，方向导数与梯度。弄清楚梯度的计算及代表意义。

沿着梯度相同方向增长最快，沿着相反方向下降最快。

# 如何使用梯度下降法

梯度下降（GD）是最小化风险函数、损失函数的一种常用方法，随机梯度下降和批量梯度下降是两种迭代求解思路.

下面的h(x)是要拟合的函数，J(theta)损失函数，theta是参数，要迭代求解的值，theta求解出来了那最终要拟合的函数h(theta)就出来了。其中m是训练集的记录条数，j是参数的个数。

**h(x)**是需要拟合的函数。

**J(θ)**称为均方误差或cost function。用来衡量训练集众的样本对线性模式的拟合程度。

**m**为训练集众样本的个数

**θ**是我们最终需要通过梯度下降法来求得的参数。

最终的目的是为了尽量缩小误差值，即求解**J(θ)的最小值。**

**J(θ)的梯度,计算结果为一个向量，即每一个θ的小梯度集合。**

**其中单个的梯度计算：**

由于要求下降值，所以要沿着梯度的反方向才是最快的下降速度

**重复上次过程，直到方向导数取值趋于收敛。**