1.最小二乘法

ref: https://zhuanlan.zhihu.com/p/38128785

优点： 直接

缺点：

1) 有可能它的逆矩阵不存在，这样就没有办法直接用最小二乘法了，此时梯度下降法仍然可以使用

2) 当样本特征，计算逆矩阵很耗时

3）如果不是线性模型，则不可用最小二乘法

1.1一元线性模型（见过知道）

考虑一元线性模型

$$h_\theta(x) = \theta_0 + \theta_1 x$$

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \theta^T X_i)^2$$

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \theta_0 - \theta_1 x_i)^2$$

链式法则

函数的极值点，为偏导数为0的点，我们有

$$\begin{cases} \frac{\partial J(\theta)}{\partial \theta_0} = 2 \cdot \frac{1}{2}\sum_{i=1}^{n}(y_i - \theta_0 - \theta_1 x_i)(-1) = 0 \\ \frac{\partial J(\theta)}{\partial \theta_1} = 2 \cdot \frac{1}{2}\sum_{i=1}^{n}(y_i - \theta_0 - \theta_1 x_i)(-x_i) = 0 \end{cases}$$

解得 $\begin{cases} \hat{\theta}_0 = \frac{\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i}{n \sum x_i^2 - (\sum x_i)^2} \\ \hat{\theta}_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \end{cases}$

对于上海房价数据（万元）：

2009年	2010年	2011年	2012年	2013年	2014年	2015年	2016年	2017年	2018年
1.8	2.1	2.3	2.3	2.85	3.0	3.3	4.9	5.45	5.0

我们归一化数据
年份为x = [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]，减去2009年
房价为y = [1.80, 2.10,2.30,2.30,2.85,3.00,3.30,4.90,5.45,5.00]，除以10000
我们通过上式计算，求得：

$$\begin{cases} \theta_0 = 1.434546 \\ \theta_1 = 0.414545 \end{cases}$$

1.2多元线性模型（熟悉结论和过程）

矩阵形式表达为：

$$min\,||A\beta - Y||_2^2 \quad A \in R^{m*(n+1)}, \beta \in R^{(n+1)*1}, Y \in R^{m*1}$$

最后求得：

$$\hat{\beta} = (A^T A)^{-1} A^T Y$$

问题：

奇异矩阵没有逆矩阵，$(A^TA)^{-1}$会出现无法求解的问题，也就是该方法对数据是有约束的，而现实中很多数据无法满足该约束

换个符号

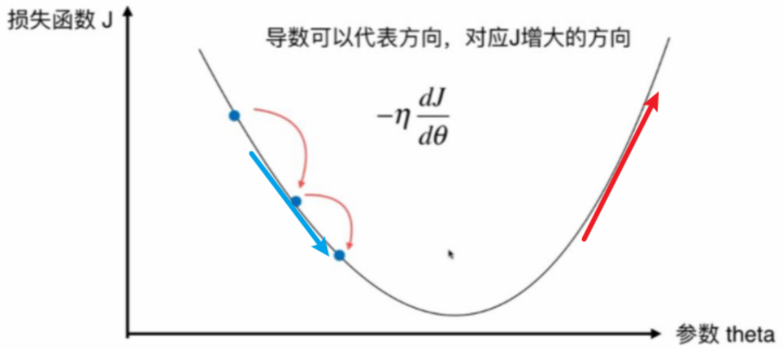
$$\theta = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

其中X【nxm】 theta 【mx1】 Y【nx1】

2. 梯度下降

2.1直观理解与导数的概念

下山 2个考虑点：步长 和 方向

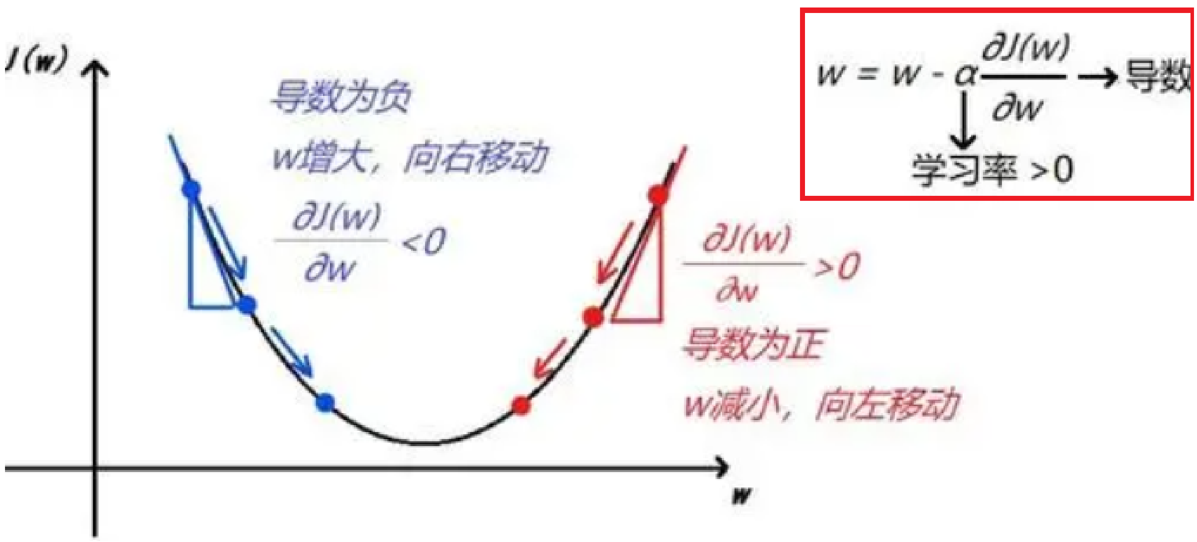


回忆一下什么是导数

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

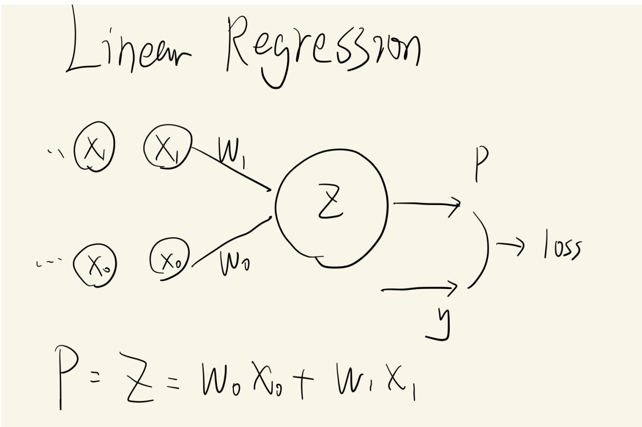
导数可以理解为是斜率

为了到达山底，我们引入了最终梯度下降的公式



特点：越到局部最小值，梯度越小，更新也越小

2.2梯度下降法的流程：



$$J(\theta) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \theta^T X_i)^2$$

在梯度下降法中，对于函数J求取极小值，是通过找到函数下降最快的方向，一步一步迭代逼近完成的。函数J被不同场合称为代价函数或者损失函数，而上式一般称之为L2损失或L2 Loss

对于函数J，我们求偏导有：

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \theta_0 - \theta_1 x_i)^2$$

$$\frac{\partial J(\theta)}{\partial \theta_0} = 2 \cdot \frac{1}{2}\sum_{i=1}^{n}(y_i - \theta_0 - \theta_1 x_i)(-1) = \sum_{i=1}^{n}(y_i - \theta_0 - \theta_1 x_i)(-1)$$

$$\frac{\partial J(\theta)}{\partial \theta_1} = 2 \cdot \frac{1}{2}\sum_{i=1}^{n}(y_i - \theta_0 - \theta_1 x_i)(-x_i) = \sum_{i=1}^{n}(y_i - \theta_0 - \theta_1 x_i)(-x_i)$$

对于求导有困难的同学可以请教一下周围学过高等数学的同学

2.3coding time

==> Linear_regression.ipynb

看矩阵求导