目录

[一、宏观 2](#_Toc28326)

[1、有监督学习（分类、回归）： 2](#_Toc32380)

[2、无监督学习（聚类）： 2](#_Toc7629)

[二、线性回归 3](#_Toc22416)

[1、单变量情况： 3](#_Toc5779)

[2、多变量情况： 3](#_Toc14929)

[3、线性回归表达式： 4](#_Toc31668)

[三、衡量线性回归的指标 7](#_Toc1258)

[1、MSE、RMSE、MAE 7](#_Toc18329)

[2、最好的评价线性回归的指标: R Squared 8](#_Toc20802)

[四、多元线性回归 9](#_Toc14615)

[五、线性回归总结 10](#_Toc17754)

[1、特点 10](#_Toc7862)

[2、最小二乘的解法： 11](#_Toc31715)

[六、参考链接 11](#_Toc15756)

**线性回归（Linear Regression）**

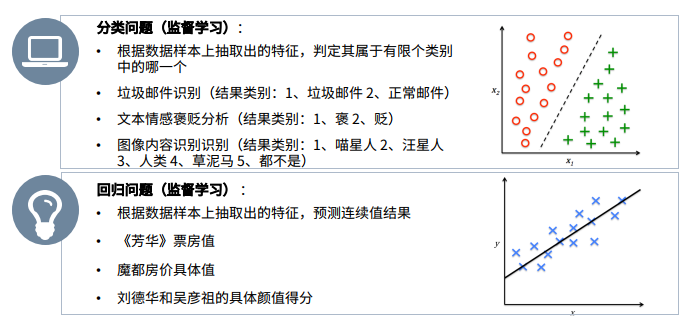
## 一、宏观

机器学习：利用大量的数据样本，使得计算机通过不断学习获得一个模型，使用该模型对新的未知数据进行预测。

**机器学习是数据驱动的算法，数据驱动=数据+模型，模型就是输入到输出的映射关系。**

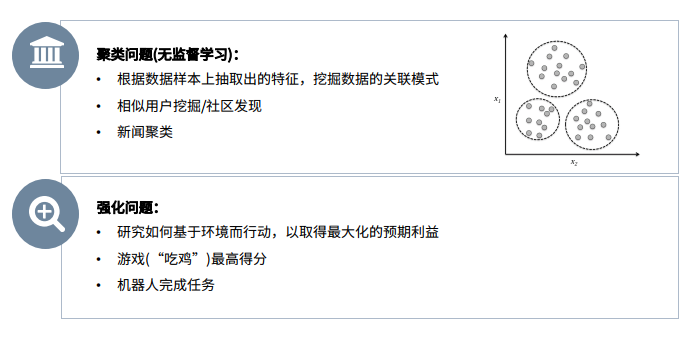
### 1、有监督学习（分类、回归）：

同时将数据样本和标签输入给模型，模型学习到数据和标签的映射关系，从而对新数据进行预测。

****

### 2、无监督学习（聚类）：

只有数据，没有标签，模型通过总结规律，从数据中挖掘信息。



强化学习：

强化学习会在没有任何标签的情况下，通过先尝试做出一些行为得到一个结果，通过这个结果是对还是错的反馈，调整之前的行为，就这样不断的调整，算法能够学习到在什么样的情况下选择什么样的行为可以得到最好的结果。

就好比你有一只还没有训练好的小狗，每当它把屋子弄乱后，就减少美味食物的数量（惩罚），每次表现不错时，就加倍美味食物的数量（奖励），那么小狗最终会学到一个知识，就是把客厅弄乱是不好的行为。

## 二、线性回归

**需要自变量与因变量之间是线性关系；**

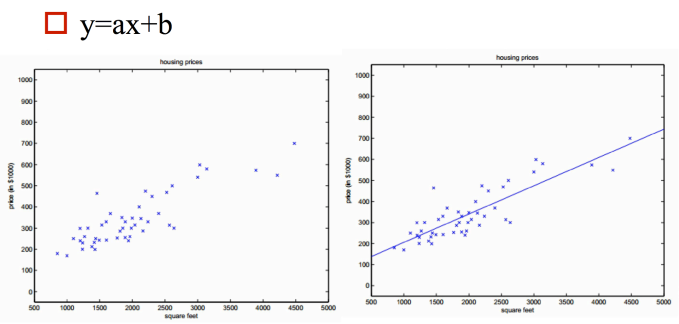
**要求误差服从正态分布；**

利用大量的样本，通过有监督的学习，学习到由x到y的映射f ，利用该映射关系对未知的数据进行预估，因为y为连续值，所以是回归问题。离散数据的则归为分类问题。

回归分析中，只包括一个自变量x和一个因变量y，且二者的关系可用一条直线近似表示，这种回归分析称为**一元线性回归**分析。如果回归分析中包括两个或两个以上的自变量，且因变量和自变量之间是线性关系，则称为**多元线性回归**分析。

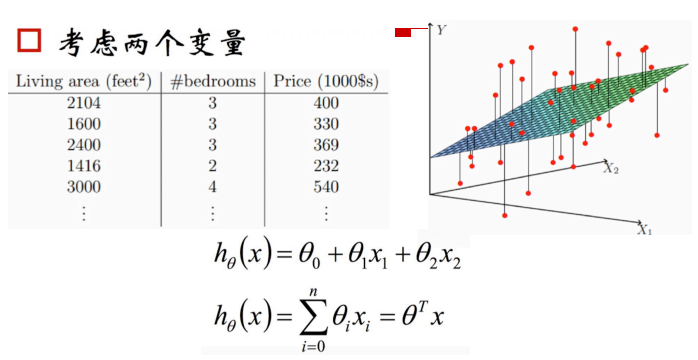
最小二乘法：拟合曲线和数据的误差平方和最小。(可以上矩阵解法，代数法计算量大)

### 1、单变量情况：



### 2、多变量情况：

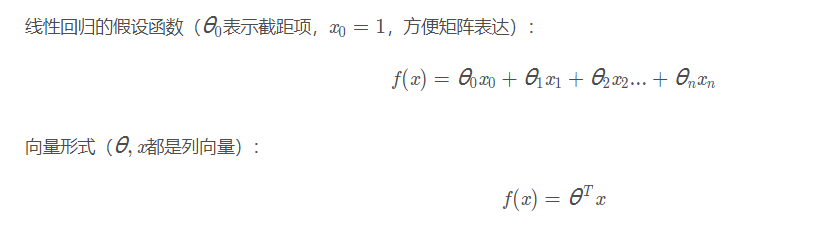
二维空间直线转化为高维空间平面



### 3、线性回归表达式：

模型=假设函数（不同的学习方式）+优化

1. **假设函数**



参数 是待求的。

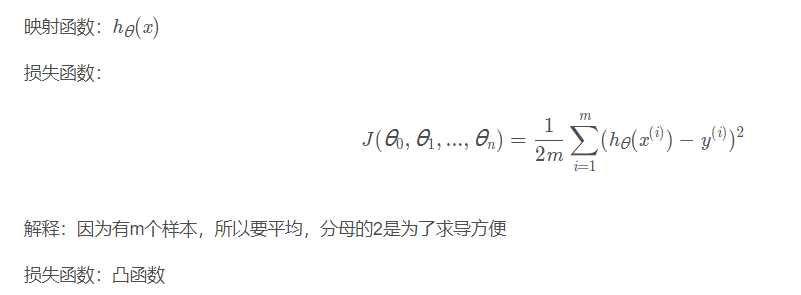
1. **优化方法**

监督学习的优化方法=损失函数+对损失函数的优化

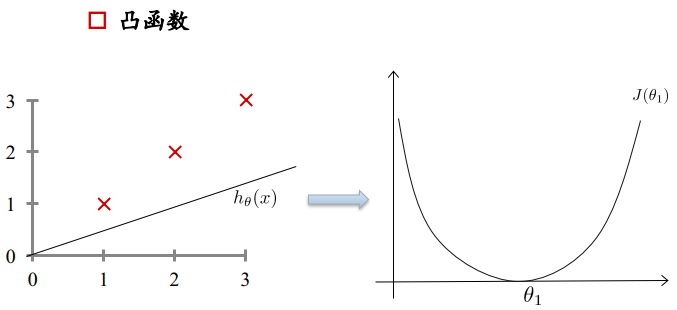
1. **损失函数**

如何评估求出的参数的好坏？

损失函数度量预测值和标准答案的偏差，通过最小化损失函数来评估优劣。



**4、损失函数的优化**

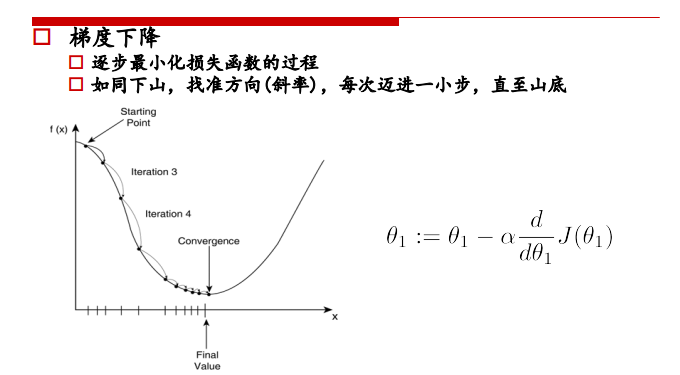


左图是需要拟合的函数h(x)，右图是损失函数J，目标是达到J的最小值点，即使得损失函数最小。

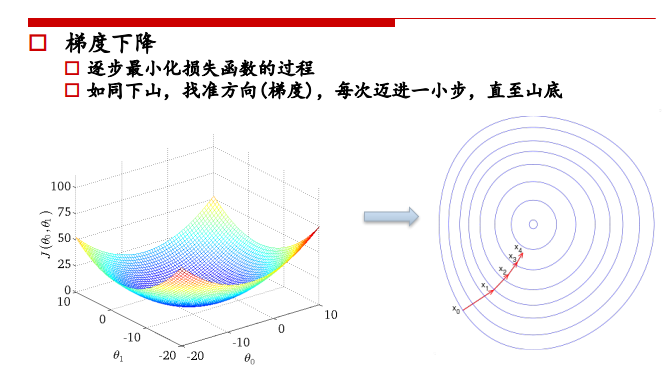
多元情况时，会出现局部极值情况。

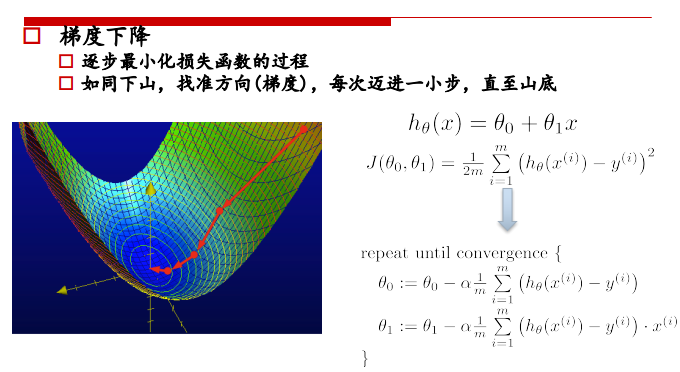
根据求极值的数学思想，对损失函数求偏导，分别解出待求参数（类比解最小二乘的代数法）。但在工业领域计算量太大，公式复杂，因此使用梯度下降法求解。

**5、梯度下降法：**



1. 初始位置选取很重要；
2. 负梯度方向更新。二维情况下，函数变换最快的方向是斜率方向。多维情况下就是梯度，梯度表示函数值增大的最快方向，所以要在负梯度方向上进行迭代；
3. 对于的更新：每个参数都是分别更新的；



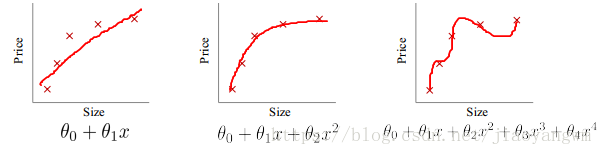


**学习率：**

学习率太大，会跳过最低点，可能不收敛；

学习率太小，导致收敛太慢；

**6、过拟合和欠拟合**



**过拟合的原因：**

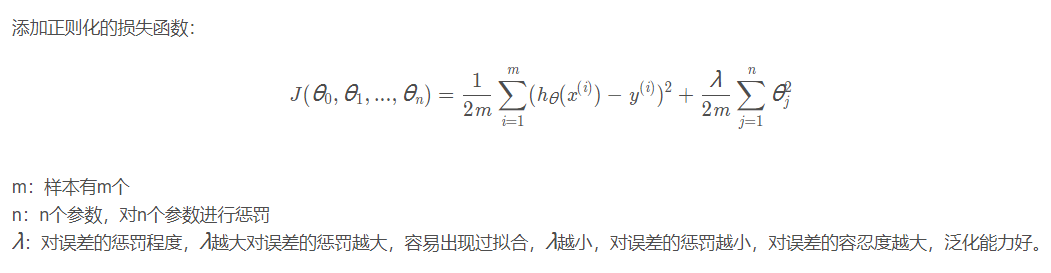
① 如果我们有很多的特征或模型很复杂，则假设函数曲线可以对训练样本拟合的非常好，学习能力太强了，但是丧失了一般性。

② 眼见不一定为实，训练样本中肯定存在噪声点，如果全都学习的话肯定会将噪声也学习进去。

**过拟合造成什么结果：**

过拟合是给参数的自由空间太大了，可以通过简单的方式让参数变化太快，并未学习到底层的规律，模型抖动太大，很不稳定，variance变大，对新数据没有泛化能力。

**正则化解决过拟合问题：**



正则化的作用：

① 控制参数变化幅度，对变化大的参数惩罚

② 限制参数搜索空间

**正则两种：**

**岭回归：**

加入L2正则

**LASSO回归：**

加入L1正则

**ElasticNet(弹性网络)：**

综合考虑岭回归和LASSO回归，按权重加入L1和L2正则得到：

1626938726(1)

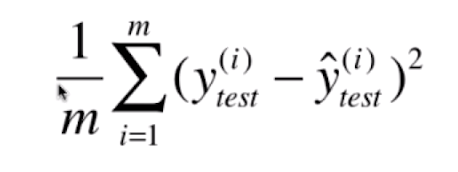
w为待求参数；

## 三、衡量线性回归的指标

### 1、MSE、RMSE、MAE

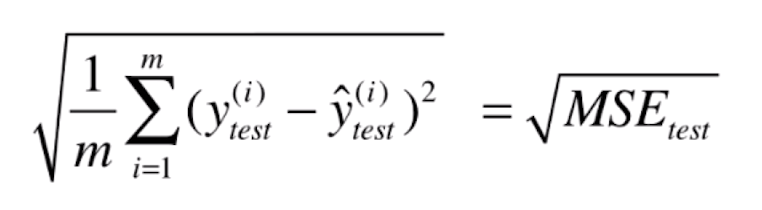
~~和测试集的规模N有关？~~分母除以N的由来。

同样使用最小二乘思想，构建均方误差MSE：



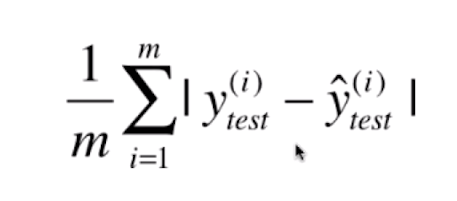
MSE量纲和y量纲不一致。

注意量纲差异--->构建均方根误差RMSE：



RMSE量纲和y量纲一致。

平均绝对误差MAE：



MAE量纲和y量纲一致。

**RMSE vs MAE:**

RMSE的结果比MAE结果大一些；

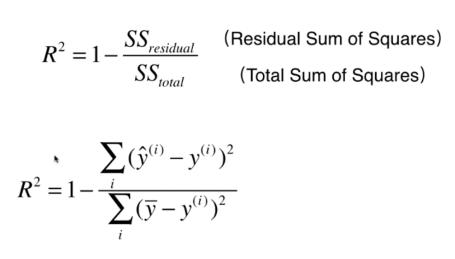
RMSE有放大样本中预测结果和真实结果之间差距的趋势，因为公式中带了平方。

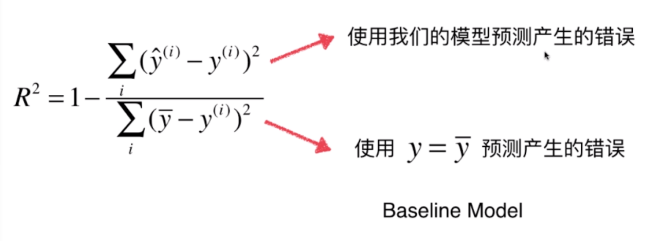
MAE没有这个趋势，直接反馈样本预测结果和真实结果之间的差距，没有平方操作。

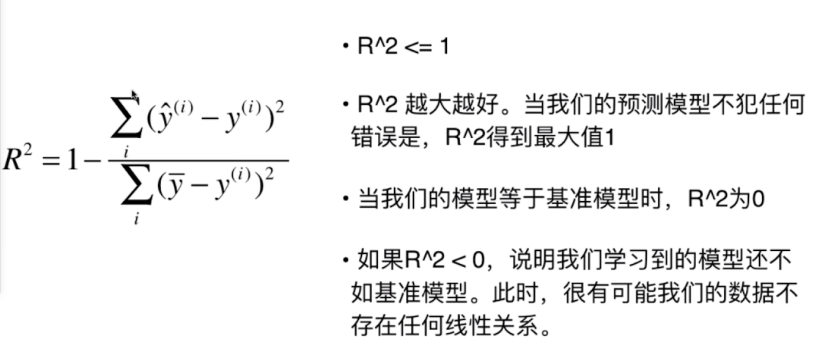
尽量让RMSE的结果更小， 相对来说意义更大一些。

### 2、最好的评价线性回归的指标: R Squared

Sk-learn中封装的算法：







解析：

（1）分子是我们训练得到的模型（和x,y有关），误差小错误少，分母把y均值当成一个基准模型，与x是无关的，误差大错误多；

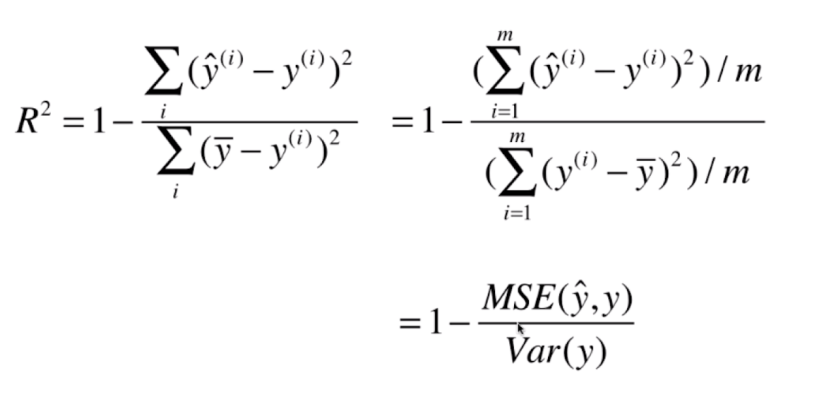
（2）把回归问题的结果拉伸到**<1**，可以将相同的回归算法应用到不同的领域；

（3）存在R2**小于0**的情况；

（4）分数是计算“犯错”的概率，整体则表示正确的概率，越大越好；

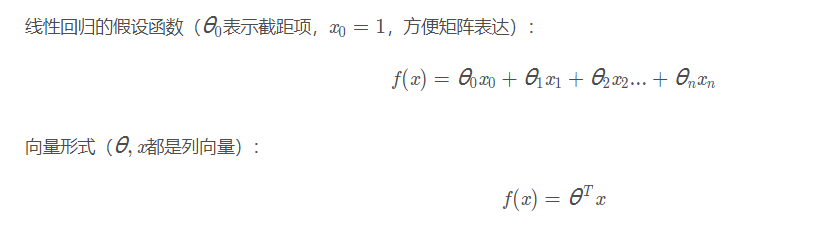
（5）若R2 = 0则表示，该组数据使用y均值模型便可以表示，那便不用费力去优化了，类似有了先验：靠近y均值的模型即可；

公式变形：



分子就是MSE，分母是方差。

## 四、多元线性回归

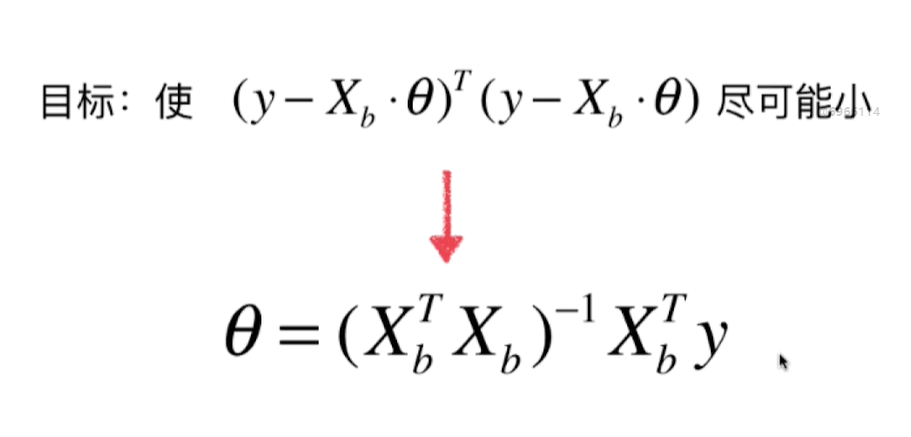


如上所示，

一元线性回归，θ0=0，当只有x1这个变量时，我们需要求解θ0和θ1两个未知参数，

类似于求解y=ax+b。

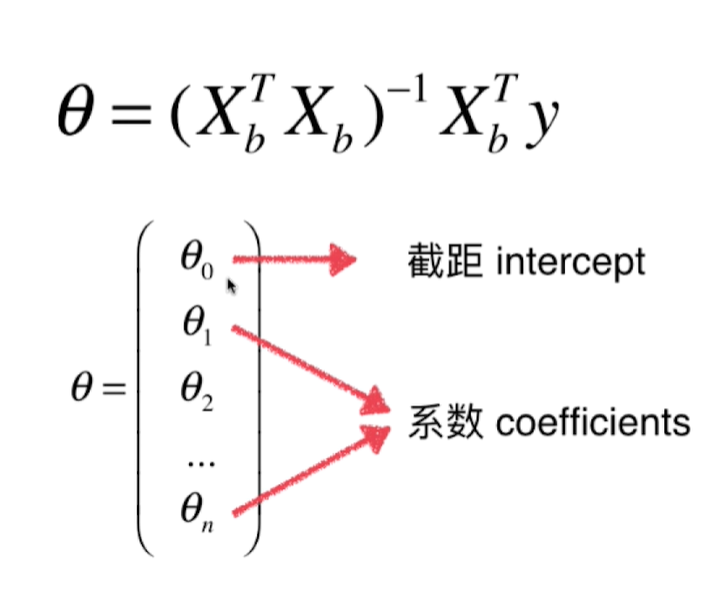
多元线性回归，包含**n**个x变量，对应需要求解(n+1)个θ参数。



对于多元回归求解，仍然是使用最小二乘法。通过矩阵转换和推导得到如上图所示的参数θ求解方程，称为多元线性回归的正规方程解（Normal Equation）。

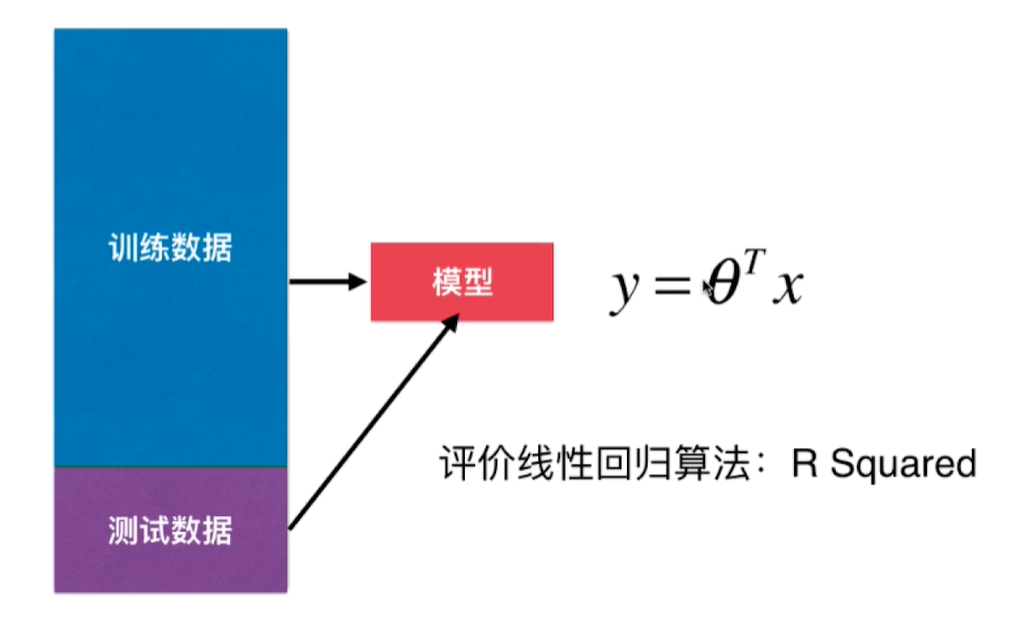
复杂度太高：大约O(N^3)；

优点：数据不用做归一化处理，无量纲问题；



得到的参数θ为向量，其包括截距和系数，每一个系数对应一个特征参数x，可以理解为，**系数表征每个特征的贡献程度**。特征参数的维度和需要求解的系数个数是相同的，再需要多求出一个截距参数即可。

## 五、线性回归总结



### 1、特点

监督学习；

参数学习；

对数据是有假设的，线性；

只能解决回归问题，在很多回归问题中，线性回归是基础（比如逻辑回归）；

线性回归的思路可以改进后去解决非线性问题；

有很强的可解释性；

### 2、最小二乘的解法：

**针对线性回归的损失函数J为最小二乘的形式，求解方法为：**

1. **代数求导**
2. **矩阵分解**
3. **梯度下降**
4. **牛顿法**
5. **高斯牛顿法**
6. **列文伯格-马夸特法**

## 六、参考链接

<https://blog.csdn.net/weixin_39881922/article/details/79915392>

好文：<https://blog.csdn.net/jiaoyangwm/article/details/81139362>

最小二乘的解法：<https://zhuanlan.zhihu.com/p/113946848>

<http://home.ustc.edu.cn/~yqli1995/2020/07/05/lr/#%E9%9D%A2%E8%AF%95%E7%9B%B8%E5%85%B3%E9%97%AE%E9%A2%98>