**机器学习**

监督学习：数据集包含正确结果

Regression回归：预测连续值的属性 Classification分类：预测离散值的输出

无监督学习：数据集不包含正确结果，仅包含数据

聚类-分簇（自动分类） 混合数据分离

**线性回归-代价函数**

拟合出的函数：

代价函数：

该函数根据预测值与样本值的最小二乘法得出

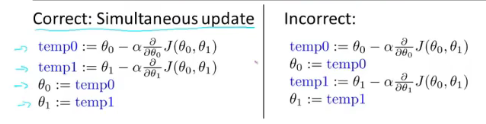
**线性回归-梯度下降**

我们要得到最佳拟合直线，就要使得代价函数的值最小，此时拟合误差最小

梯度下降是通过对代价函数求偏导并修正于原本的位置上，使得代价函数逐渐趋向于局部最小值。当其达到局部最小值以后，梯度下降将不会对系统参数产生影响。同时，在多维代价函数趋势较为复杂时，其可能有多个局部最小值，因此当梯度下降的起点不同时，其结果也可能不相同。在线性回归中，代价函数仅有一个局部最小值，其同时也是全局最小值。更加复杂的情况以后再进行讨论。

线性回归梯度下降算法的具体实现过程：

 该式计算两次，同时计算修正拟合曲线的两个参数后再赋值



此式子中的α为学习率，其取值如果过小，则需要很多次迭代才能达到目标效果，效率较低。若取值过大，则可能导致代价函数经过修正后发散，完全无法取得结果。

另外，遍历所有样本的梯度下降叫做batch梯度下降。

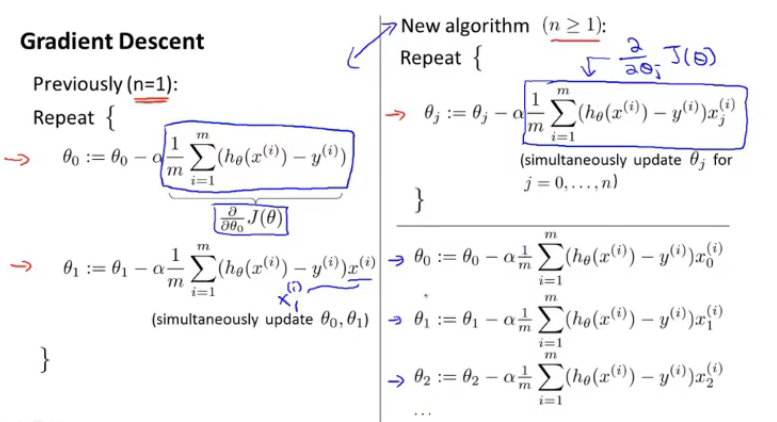
**多元线性回归**



（X0一般取为1。第0项表示常数修正）同时预测等于系数矩阵的转置乘特征量矩阵

**多元线性回归-梯度下降**

其原理相同，但是因为函数变为多元，其具体求值表达式也不相同



**多元线性回归-特征缩放**

当几个特征值的范围之间相差好几个数量级的时候会导致损失函数的形状过于偏向于一个方向，因此在进行梯度下降时很容易发生震荡现象，导致修正效果较差。

通常使用来进行特征值的标准化。此处μ为特征值的平均值，s是特征值包括的范围。实际上当特征值的区间相近即可，不需要保证其高度一致性。

**多元线性回归-梯度下降学习率调试**

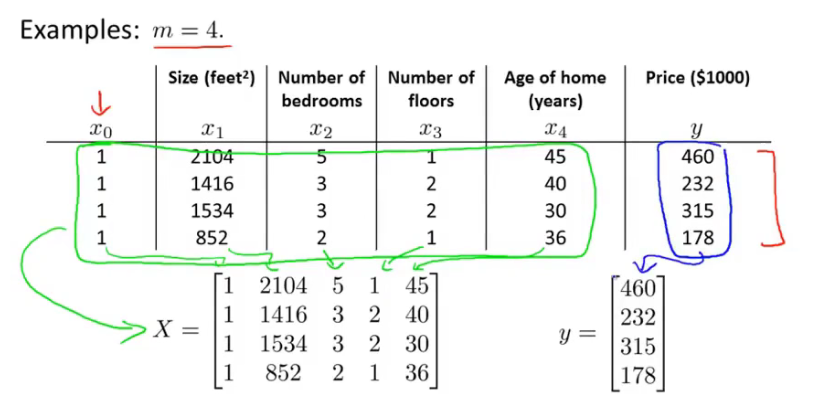
可以绘制关于迭代次数的函数图，观察随着迭代次数的增加，代价函数的下降情况。通常在一次迭代后代价函数的下降值小于时，认为此时代价函数已经收敛。当下降过于缓慢时，增加学习率。当代价函数上升或者先下降后上升，减小学习率。测试时可以按照0.003，0.01，0.03这样的规律来测试。

**多项式回归-特征**

如果线性回归无法很好的拟合数据集，我们可以构造多项式回归来进行更好的拟合。我们也可以根据数据集里的特征自己合成或分解出新的特征来使用。同时在进行多项式回归时，特征缩放尤其重要。

**多项式回归-正规方程求代价函数最小值**

按照数据集构造特征矩阵和结果矩阵



求解表达式：

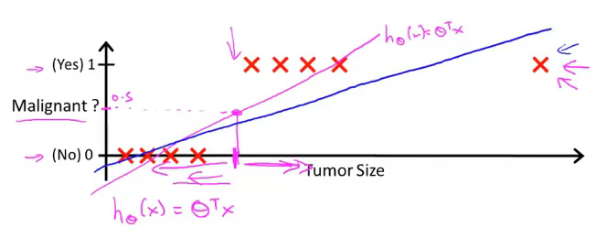
当不可逆的时候，通常是因为有多余的特征或者特征数量远远超出样本数量。（从线性代数角度来说，矩阵不可逆通常是因为其行列式等于0或者其非满秩矩阵）此时，我们应该试图删除一些重复的或者不重要的特征以减少特征种类。实际上，对不可逆矩阵直接求伪逆也可以得到正确结果。同时正规方程求解法不需要进行特征缩放。

**多项式回归-梯度下降与方程求解的选择**

方程求解不需要经过迭代，且不需要选择学习率α。在大多数情况下，此方法都是更优的求解方法。但是由于方程求解需要运行这一步，其时间复杂度为特征个数n的三次方。在特征个数非常多的时候，方程求解运算速度会显著下降。在实际应用时，特征个数n大于10000时我们就倾向于使用梯度下降法。

**逻辑回归-分类**

分类问题也可以使用线性回归，但是一旦出现极端点，会导致其拟合的直线效果很差。



以0.5作为分界，当出现极端值时样本集中区域会出现大量的判断错误现象，因此需要使用一定手段避免极端值的影响。

**逻辑回归-假设函数**

假设函数即为算法拟合出的函数。逻辑回归的假设函数为：

此处即为一般的多项式回归值。为非线性化函数，

此处的称为决策边界，此函数将样本划分在两个不同的区域中。

**逻辑回归-代价函数**

引入概念-cost函数：

对线性回归来说，=

使得 

如果对逻辑回归仍旧使用线性回归的cost函数，会导致代价函数非凸，有多个局部最小值，因此对逻辑回归有专门的cost函数：（仅在y只取0或1时成立）



该代价函数的优势在于当判断正确时，其给出的代价惩罚极小，即趋近于0。但随着判断失准程度增大，代价惩罚快速增加，当判断完全错误时代价函数趋近于无穷。

**逻辑回归-梯度下降**

仍旧是重复梯度下降的更新流程。



此时梯度下降计算式的形式虽然看似与多元线性回归相同，但是注意表达式内涵不同。对多元线性回归，而此处

另外，此处使用梯度下降也需要注意特征缩放的必要性

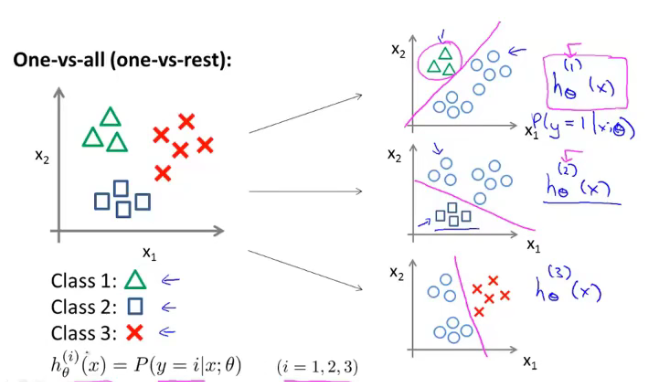
**逻辑回归-高级优化方法**

conjudate gradient（共轭梯度法）、BFGS、L-BFGS

这些算法不需要调节学习率且效率远高于梯度下降，但是原理较为复杂，此处按下不表

**逻辑回归-多分类问题**

以上的分析全部建立在二分类问题的基础上，实际上逻辑回归可以用于解决多分类问题。



我们将数据集处理后，将其中一种独立出来，其他所有视为同一种类。用处理后的伪数据集训练出多个分类器，分别用于得出一个样本属于三个不同种类的概率。在最后输出前，取置信度最高的一个种类作为逻辑回归分类器的输出。