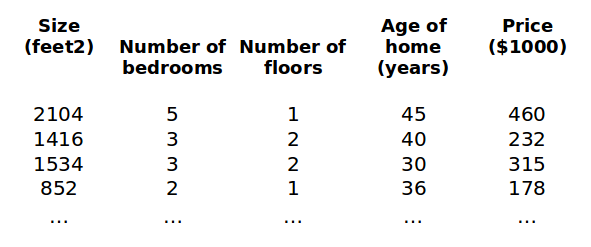
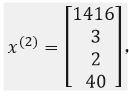
# 多变量线性回归

## 多维特征

讨论完单变量线性回归，现在我们对房价模型增加更多特征，例如房间数楼层，构成多变量模型，模型特征为

如图所示，

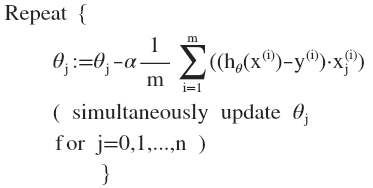


增加特征后，我们引入一系列新的注释，n代表特征的数量，代表第i个训练实例，用来是个列向量，由各个特征组成，如表示第i个样本的第j个特征

此使的假设函数变为，但如此表示并不方便，为此我们为样本额外增加特征，引入，这样假设函数便可表示为，经过向量化，公式可简化为

## 多变量梯度下降

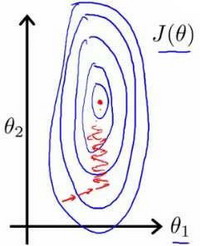
与单变量线性回归类似，我们构建代价函数，，我们要通过梯度下降来找到一系列θ值，使得J最小。

利用梯度下降算法，可得公式，

经过多轮循环，我们可得到θ的最优解

## 特征缩放

有时我们样本的特征之间差距过大，这会造成损失函数的图像难以收敛，如图所示

图像变得十分“扁平”，我们需要经过多轮迭代才能得到最优解

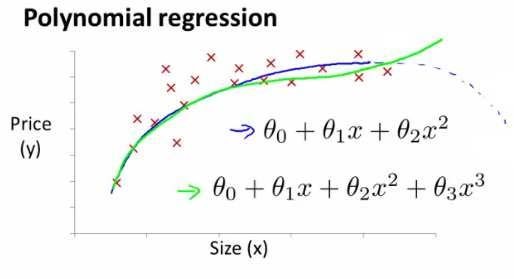
那么我们该如何解决这个问题呢？联想到几何中的仿射变换，我们需要对特征进行缩放。

让特征均位于（-1，1）之间，公式为，其中是平均值，是标准差，我们可以让新特征的平均值为0.至于为什么使用这个公式，我会在其他文章中谈到。

## 学习率

运用梯度下降算法时，我们要对学习率进行精心选择，学习率过高会导致不收敛，学习率过低会导致收敛速度过慢，占用太多的迭代时间。我们通常可以采用这些学习率进行测试，，当然，也有指数下降学习率的方法，我们之后会提到，其实这个方法并不是特别必要。

## 5．特征和多项式回归

有时用直线拟合一系列样本并不是个好主意，例如房价预测问题，一个样本有特征x1,x2分别表示房子的长和宽，如此，我们用线性回归并不能得到满意的解答，我们需要做的是增加特征x3，令x3 = x1\*x2，以此进行曲线的拟合。

我们可以通过观察样本的分布，来决定是要进行多项式拟合抑或是直线拟合

如是我们要进行多项式回归，则务必要进行特征缩放，否则比如说计算1000^5将是很可怕的事情

## 正规方程

除了运用梯度下降的方法，我们也可以用线性代数的方法直接解出最优的θ值，该方法被称为正规方程，在此不进行赘述。

公式为，当x的特征较小时，可以采用此方法，特征数量小于1w时即可考虑

该方法仅适用于线性回归模型，而不适用于其他回归，梯度下降适用性更广