

以预测房价为例

## Housing prices prediction

$$h_{\theta}(x) = \theta_0 + \theta_1 \times \text{frontage} + \theta_2 \times \text{depth}$$



假设有两个特征，分别是房子临街的宽度和垂直宽度，上图是想要卖出的房子的图片，临界宽度被定义房子的宽度，或者说是拥有土地的宽度房子的纵向深度就是房子的深度，两个称为临界宽度和纵深。建立回归模型，临街宽度是第一个特征  $x_1$ ，纵深是第二个特征  $x_2$ 。要找到决定房子价格的因素是什么？因此需要创造一个新的特征，称为  $x$  是临界宽度与纵深的乘积，得到拥有土地的面积。然后把假设使其只使用一个特征也就是土地面积来得到房子的价格即面积=长\*宽，但是实际上加上选择合理的其他的因素会得到一个更好的模型。

## Housing prices prediction

$$h_{\theta}(x) = \theta_0 + \theta_1 \times \underbrace{\text{frontage}}_{x_1} + \theta_2 \times \underbrace{\text{depth}}_{x_2}$$

Area

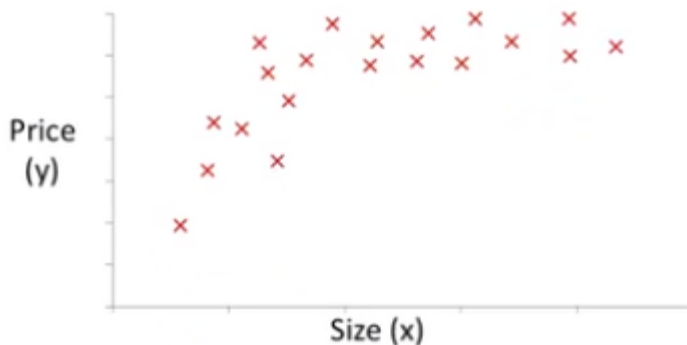
$$x = \text{frontage} * \text{depth}$$

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



多项式回归

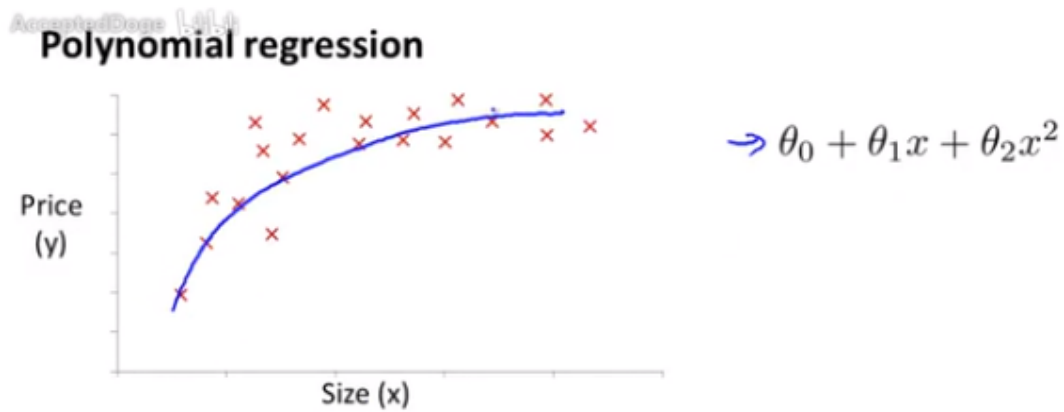
## Polynomial regression



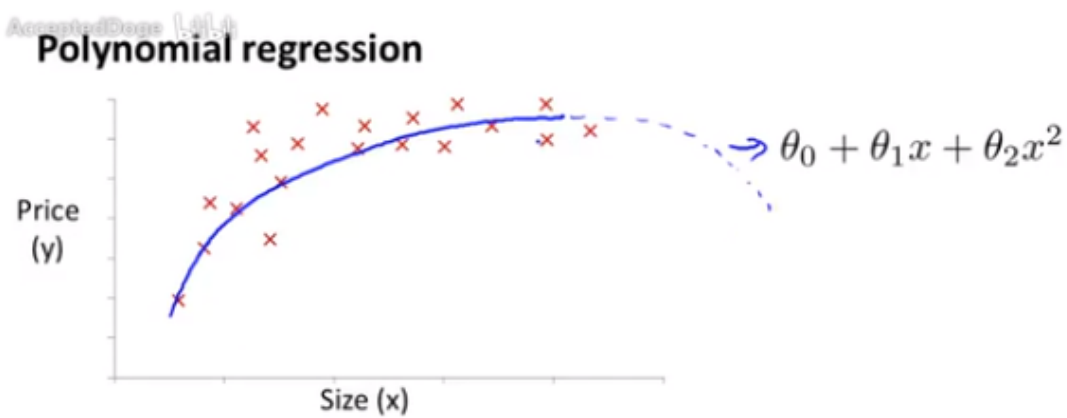
比如有一个住房价格的数据集，为了拟合它，可能会有多个不同模型可供选择，其中一个可以选择像这样的二次模型

$$\theta_0 + \theta_1 x + \theta_2 x^2$$

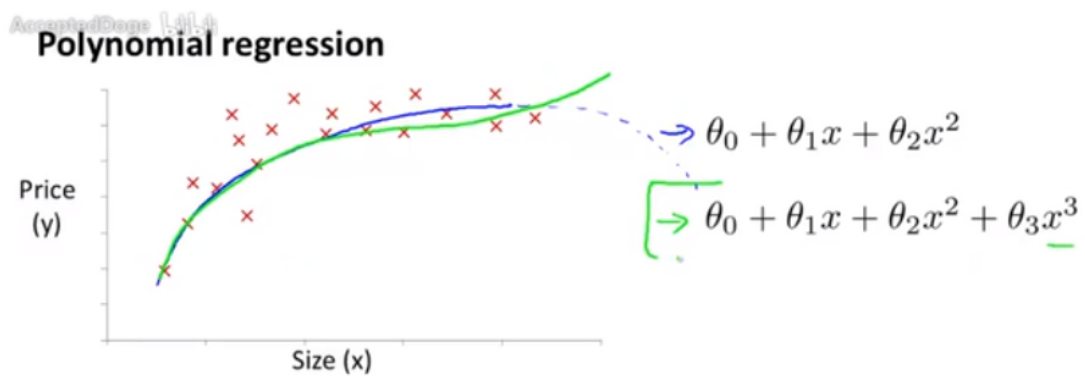
因为直线不能很好的拟合这些数据，二次模型拟合结果



但是二次模型可能不是很好，二次函数最终会降回来

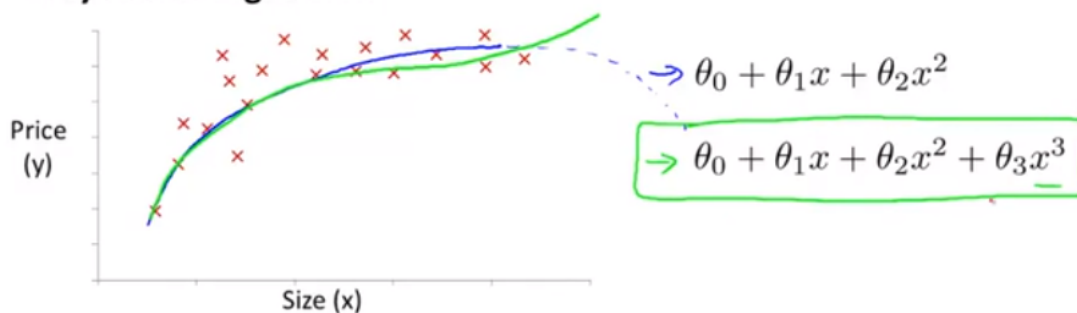


但我们并不认为，房价升到一定程度会降回来。因此，我们会选择一个不同的多项式模型并转而使用一个三次函数在这里（23.4-5 3: 05）用它来拟合这跟绿色的线



拟合的方法：使用多元线性回归的方法拟合这个三次模型

## Polynomial regression



$$h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$

$$= \theta_0 + \theta_1(\text{size}) + \theta_2(\text{size})^2 + \theta_3(\text{size})^3$$

$$x_1 = (\text{size})$$

$$x_2 = (\text{size})^2$$

$$x_3 = (\text{size})^3$$

fit this cubic model that  
拟合这个三次模型

为了预测一个房子的价格用  $\theta_0 + \theta_1 \cdot \text{size}(\text{房子面积}) + \theta_2 \cdot \text{size} \cdot \text{size}(\text{房子面积的平方}) + \theta_3 \cdot \text{size} \cdot \text{size} \cdot \text{size}(\text{房子面积的立方})$  使用  $x_1, x_2, x_3$ ，设为面积，面积的平方，面积的立方。

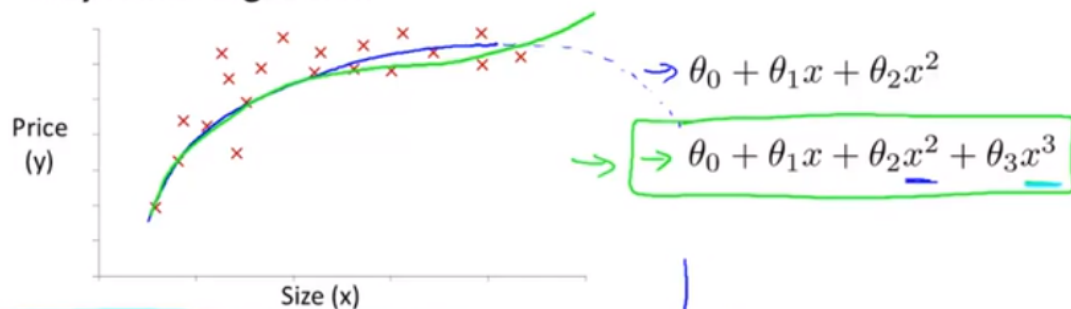
$$x_1 = (\text{size})$$

$$x_2 = (\text{size})^2$$

$$x_3 = (\text{size})^3$$

通过这样的设置在应用线性回归的方法可以拟合这个模型，最后将这个三次函数拟合到我的数据上

## Polynomial regression



$$h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3$$

$$= \theta_0 + \theta_1(\text{size}) + \theta_2(\text{size})^2 + \theta_3(\text{size})^3$$

$$\rightarrow x_1 = (\text{size})$$

$$\rightarrow x_2 = (\text{size})^2$$

$$\rightarrow x_3 = (\text{size})^3$$

like this, then feature scaling  
那么特征的归一化

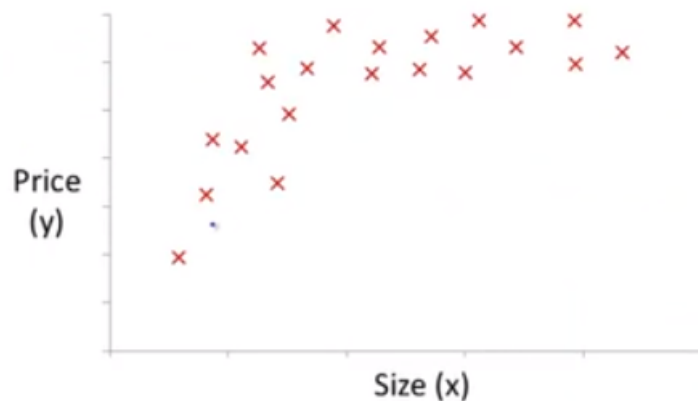
假设房子的面积是 1-1000，那么面积的平方立方为：

Size: 1-1000  
 Size<sup>2</sup>: 1-1,000,000  
 Size<sup>3</sup>: 1-10<sup>9</sup>

三个特征值的范围有很大的不同，使用梯度下降应用特征归一化是非常重要的。要将他们的值的范围具有可比性。

Example: 选出真正要使用的特征

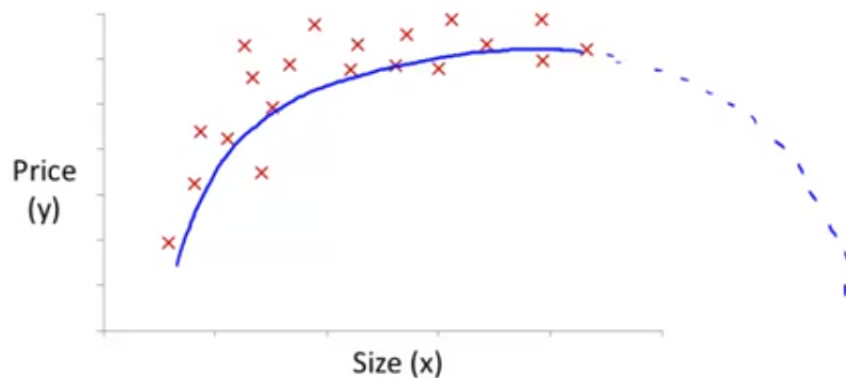
### Choice of features



$$\rightarrow h_{\theta}(x) = \theta_0 + \theta_1(\text{size}) + \theta_2(\text{size})^2$$

使用二次模型并不理想，因为二次函数最后会下降，就是住房价格往下走，像预测那样出现房价下降，除了使用三次模型以外，还可以选择平方根

### Choice of features



$$\rightarrow h_{\theta}(x) = \theta_0 + \theta_1(\text{size}) + \theta_2(\text{size})^2$$

two times the square root of the size, right?  
 加  $\theta_2$  乘以房子面积的平方根 可以吧?

这个趋势是上升的，但慢慢地去想平缓，永远不会下降回来。