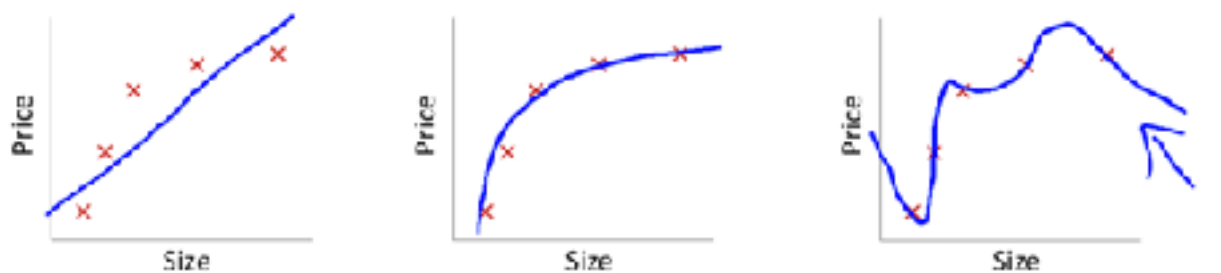


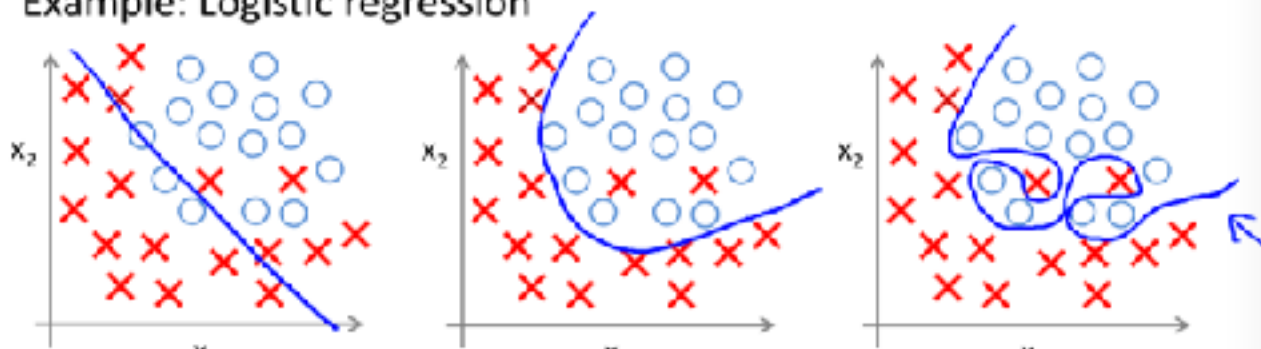
过拟合：

以多元函数回归为例子，当参数较多的时候我们可以比较好地拟合训练数据。但正是因此我们过好地拟合了训练数据，甚至好到将误差(或者叫噪声)都拟合上了，导致模型在预测新的数据的时候效果不好。从计算的角度出发是因为我们的特征太多了(相当于其中有噪声)，从而导致了过拟合的问题。如下图所示：

Example: Linear regression (housing prices)



Example: Logistic regression



解决过拟合的常用方法：

- 1.减少特征的数量：人为的手段或者是通过一些算法选择的手段
- 2.正则化：保持所有特征，但加大对较大参数的限制

正则化的通用表示：

$$J(\theta) = \frac{1}{2m} \left[\sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^n \theta_j^2 \right]$$

正则化项的计算：

Gradient descent

Repeat {

$$\rightarrow \theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_0^{(i)}$$

$$\rightarrow \theta_j := \theta_j - \alpha \left[\frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)} + \frac{\lambda}{m} \theta_j \right]$$

$j = \text{~~0~~ } 1, 2, 3, \dots, n)$
 $\theta_1, \dots, \theta_n$

}