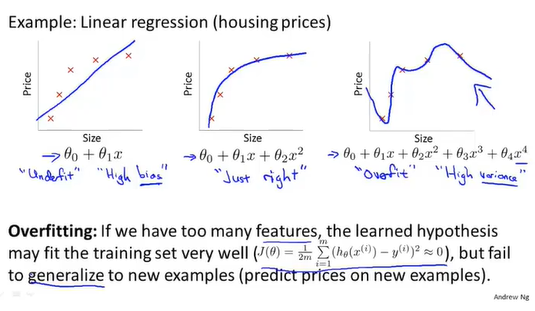
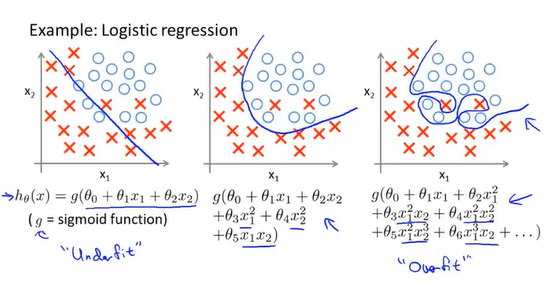
# 过拟合问题



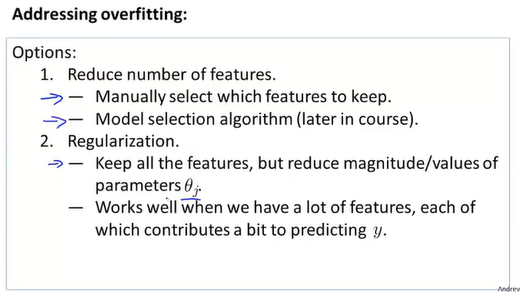
线性回归中的过拟合问题

过拟合由于变量过多，过度适合训练集J(θ)几乎等于0，所以导致它无法泛化到新样本中。



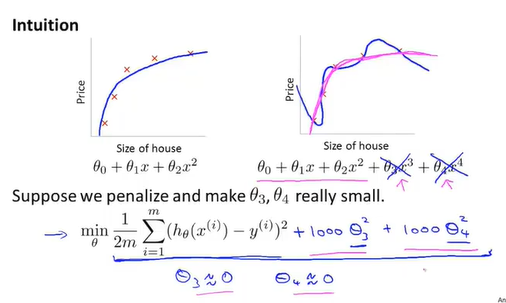
逻辑回归中的过拟合问题

# 处理过拟合问题

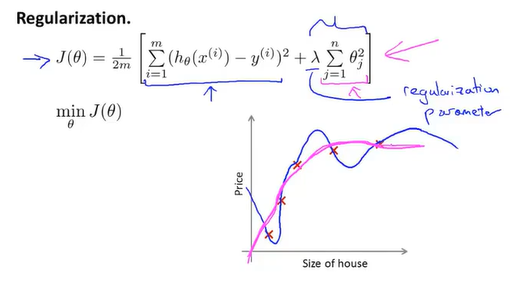
过多的变量和过少的训练数据就会导致过拟合问题  


两种方法：一是人工或者通过模型减少特征的数量，二是正则化

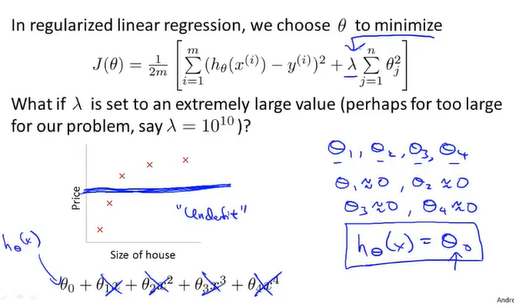
# 正则化代价函数Regularization Cost function



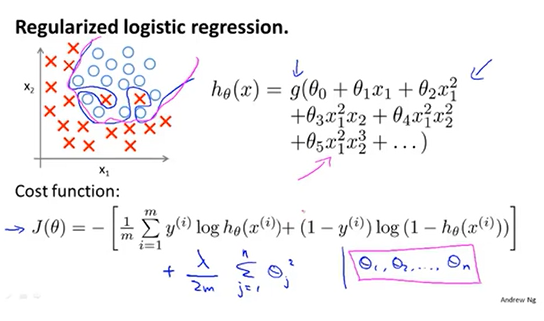
正则化思想：在cost function后面添加一个惩罚项，使得θ3和θ4趋近于0，得到一个类似于二次函数，但是又包含x3, x4这两项的函数



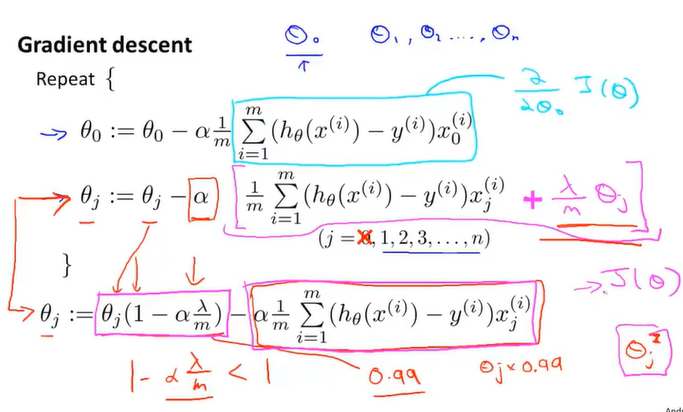
正则化代价函数



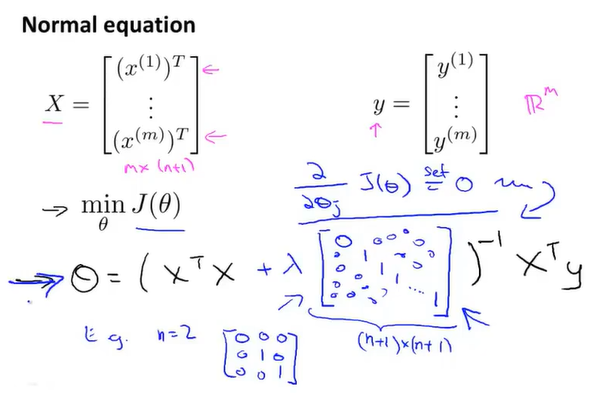
如果λ（正则化参数）很大就会导致hθ(x)=θ0，使cost function变成一条直线



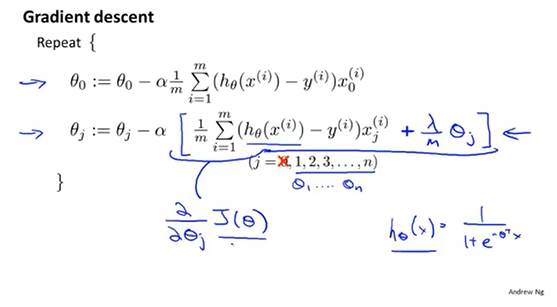
正则化应用于logistic regression（分类算法）



添加正则化项之后偏导数的变化



矩阵形式（正规方程）



正则logistic regression求导，和线性回归公式一样