评估假设：

一个训练误差最小的假设有时并不是一个好的假设，而且当特征过多时，我们很难画出假设函数来观察。

有一种标准方法，将训练数据分成两部分，第一部分是训练集，第二部分是测试集，一般是7:3左右。

具体步骤：

那么我们训练模型的流程可以有如下两个部分：

（1）使用训练集来训练模型参数θ，（最小化代价函数​Jtrain(Θ)​）

（2）使用测试集来计算误差​Jtest(Θ)​

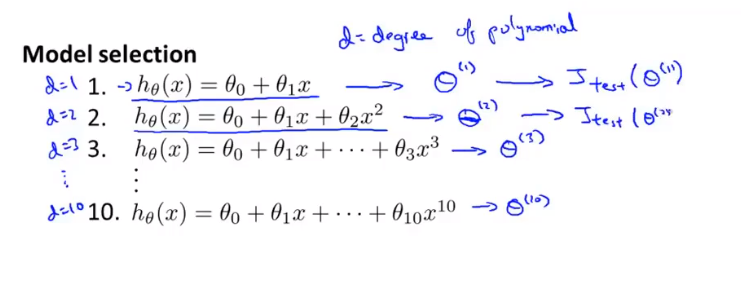
对于线性回归来说，误差：

clipboard.png

对于逻辑回归来说，误差：

clipboard.png

模型选择，交叉验证集 =》 选择d次方的多项式



如果还是使用7:3的数据划分，会有数据集重复使用问题（使用测试集计算θ，又用测试集计算d，不利于模型的泛化）

所以一般使用6:2:2的数据集划分方式

确认θ 使用 训练数据集60%

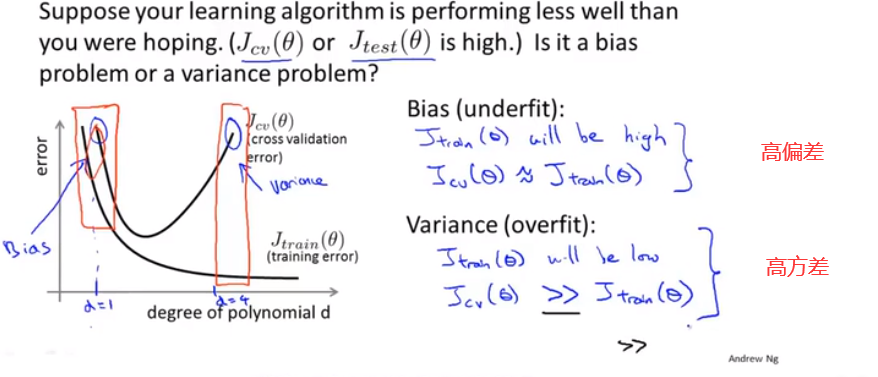
确认d 使用 验证数据集20%

测试模型 使用测试数据集20%

方差偏差 =》 选择d

高偏差欠拟合

高方差过拟合



正则化方差偏差 =》选择**λ**

