Bias and Variance

## Regularization

**Underfitting/High bias**: Our model is not complex enough to capture the pattern in the training data well and therefore also suffers from low performance on unseen data.

**Overfitting/High variance**: If we have too many features, the learned hypothesis may fit the training set very well (J() ~= 0),but fail to generalize to new examples (predict prices on new examples).

Cause: 1. Too many parameters that lead to a model that is too complex given the underlying data.

2.

**addressing overfitting**:

1. Reduce number of features

    --- Manually select which features to keep

    --- Model selection algorithm (later in course)

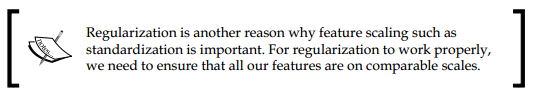
1. Regularization

--- keep all the features, but reduce magnitude/values of parameters C:\Users\phenix\AppData\Local\Temp\enhtmlclip\Image(20).png

--- Works well when we have a lot of features, each of which contributes a bit to predicting y.

--- Regularization is a very useful method to handle collinearity (high correlation among features), filter out noise from data, and eventually prevent overfitting.

L2 regularization: 



**Cost Function:**

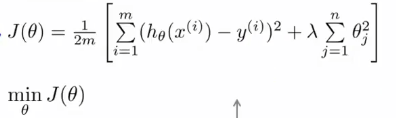
Suppose we penalize and make C:\Users\phenix\AppData\Local\Temp\enhtmlclip\Image(21).png really small.

C:\Users\phenix\AppData\Local\Temp\enhtmlclip\Image(22).png

Small values for parameters,C:\Users\phenix\AppData\Local\Temp\enhtmlclip\Image(23).png

    ---"Simplier " hypothesis

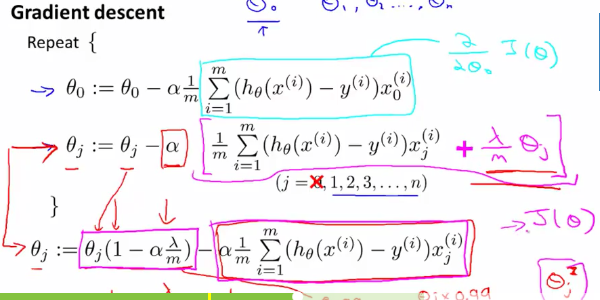
    --- Less prone to overfitting

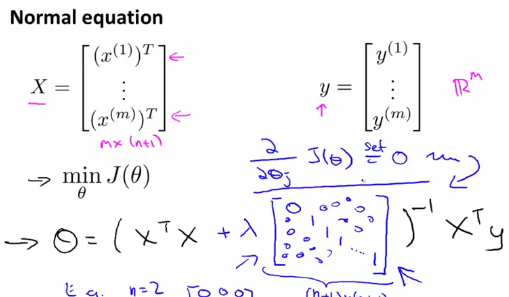


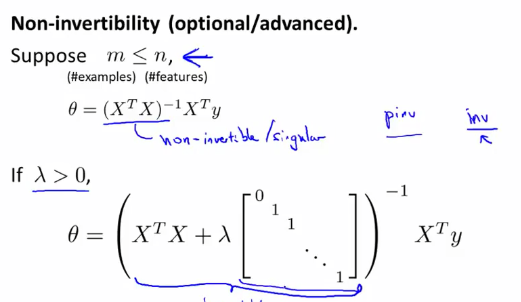
lamda is **regularization parameter**

By increasing the value of lamda, we increase the regularization strength, and the weight coefficient shrink.

**Regularized Linear Regression**







### Advice for Applying Machine Learning

#### Debugging a learning algorithm

  -- Get more training examples

  -- Try smaller sets of features

  -- Try getting additional features

  -- Try adding polynomial features

  -- Try descreasing lambda

  -- Try increasing lambda

Machine learning diagnostic:

Diagnostic: A test that you can run to gain insight what is/isn't working with a learning algorithm, and gain guidance as to how best to improve its performance.

Diagnostics can take time to implement, but doing so can be a very good use of your time.

**Evaluating a Hypothesis**

**在研究对比不同算法的泛化性能是,我们用测试集上的判别效果来评估模型在实际使用时的泛化能力,而把训练数据另外划分为训练集和验证集,基于验证集上的性能来进行模型选择和调参.**

70% training set ,30% test set

产生训练集和测试集的方法:

2.2.1 留出法（hold-out）

直接划分成互斥的集合.

需注意:

1. 训练/测试集的划分要尽可能保持数据分布的一致性,避免因数据划分过程引入额外的偏差而对最终结果产生影响.保留类别比例的采样方式通常称为”分层采样”(stratified sampling)

一般采用若干次随机划分、重复进行实验评估后取平均值作为留出发的评估结果

  Trianing/testing proceduce for linear regression

 -- Learning parameter C:\Users\phenix\AppData\Local\Temp\enhtmlclip\Image(31).png from training data (minimizing training error J ())

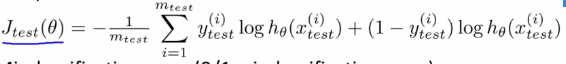
--compute test set error:



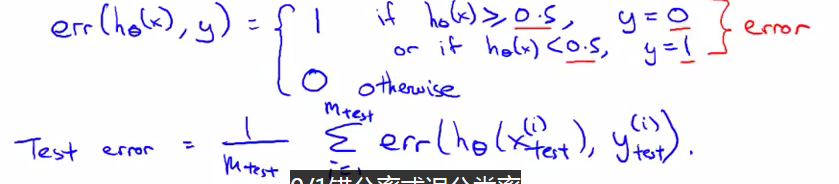
Training/testing proceduce for logistic regression

 -- Learning  from training data

  -- Compute test set error:



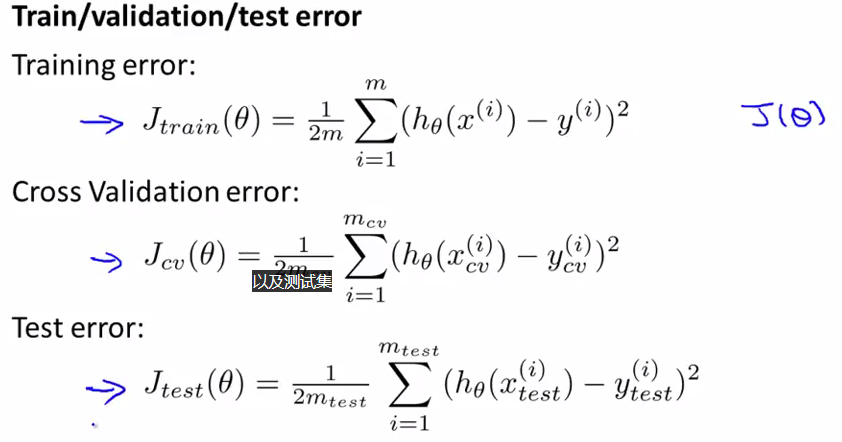
-- Misclassification error (0/1 misclassification error):

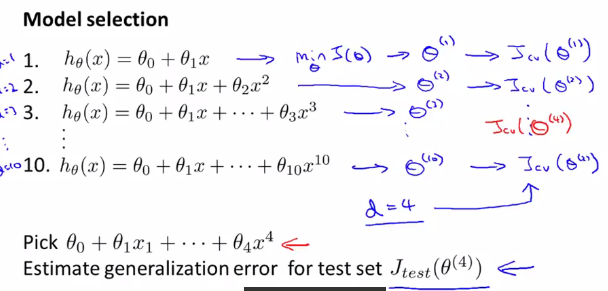


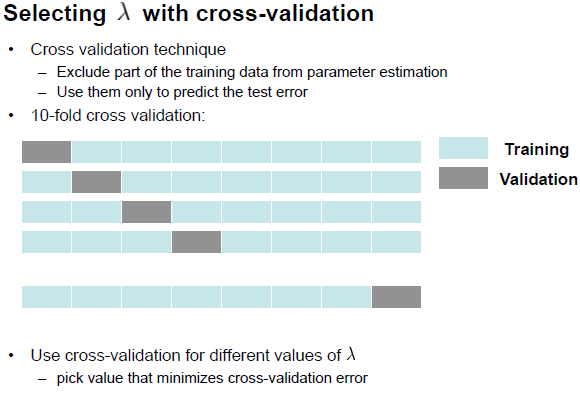
**Model Selection and Training/validation/Test Sets**

交叉验证

60% Training Set, 20% cross Validation set , 20% test set





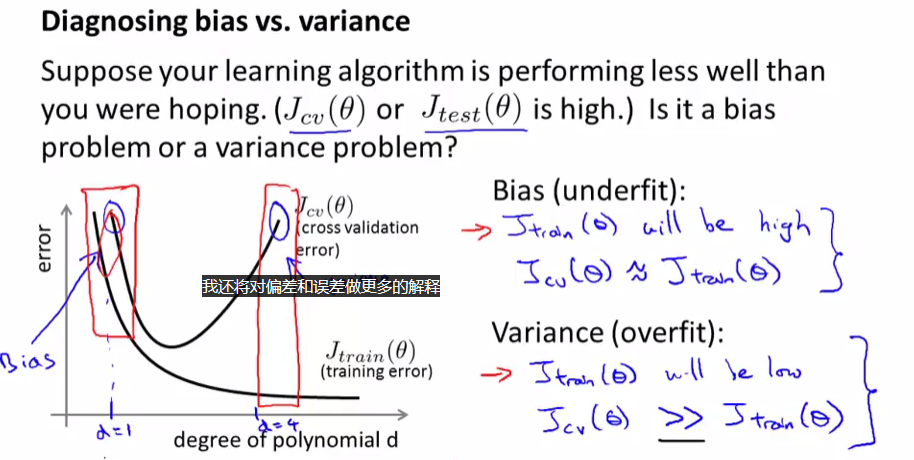
 Bootstrapping(自助法):

有放回采样.

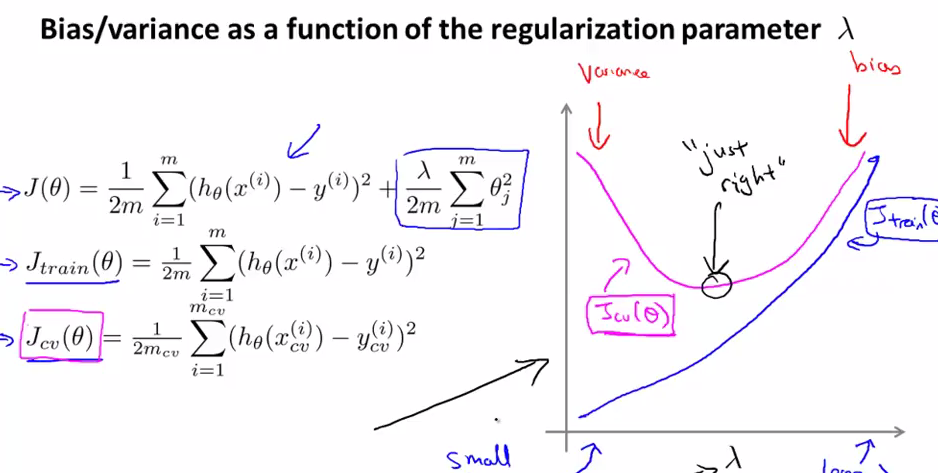
Diagnosing Bias vs Variance

偏差-方差分解:

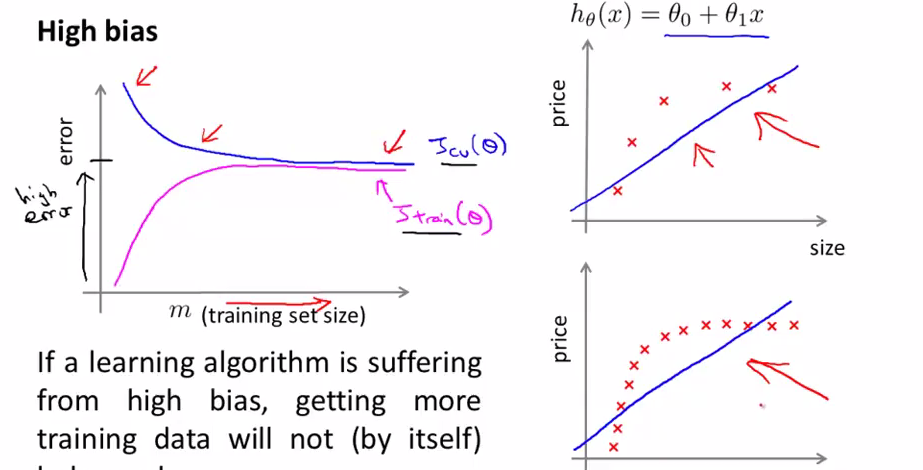
对测试样本x,另为想在

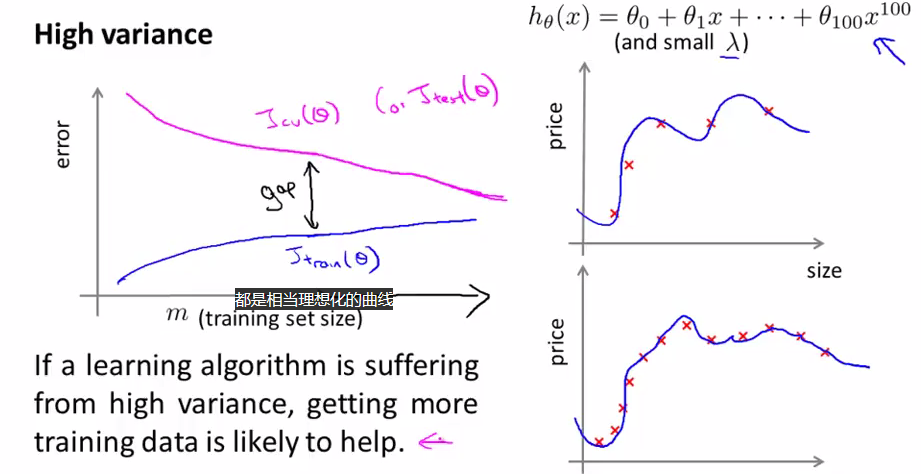


Regularization and Bias/Variance

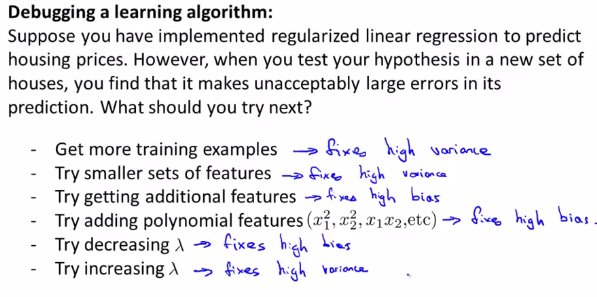


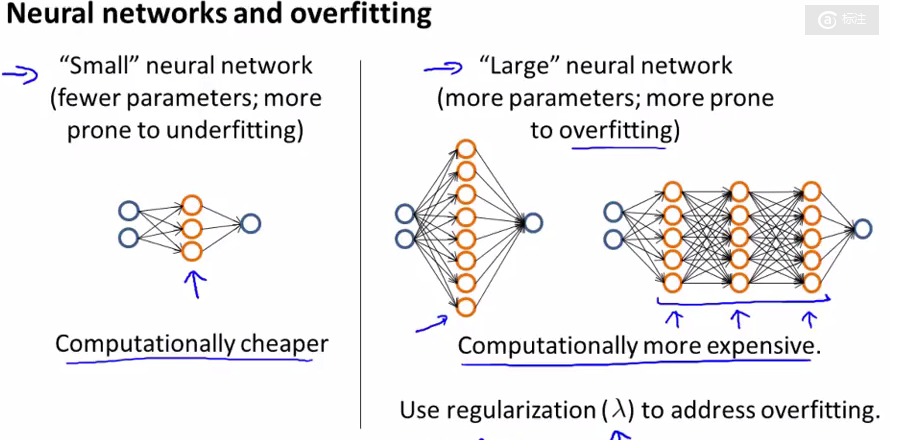
Learning Curves





Deciding What to Do Next Revisited





Feature scaling(特征缩放)

     Idea: Make sure features are on a similar scale.

     Get every feature into approximately a -1<= xi<=1 range

Mean normalization（规范化）

     Replace xi with xi-ui to make features have approximately zero mean (Do no apply to x0 =1

Reference:

1. Machine learning , Ng video.
2. Python machine learning, book.
3. 机器学习-周志华