# 开始整理自学部分的知识框架？

# 不知道能不能坚持

主要分类方式：

1. supervised learning

we are told what is suppose to be the right answer

1. unsupervised learning

只要所给的数据的输入输出关系不明确，哪怕确定只有两种输出，仍然视为2

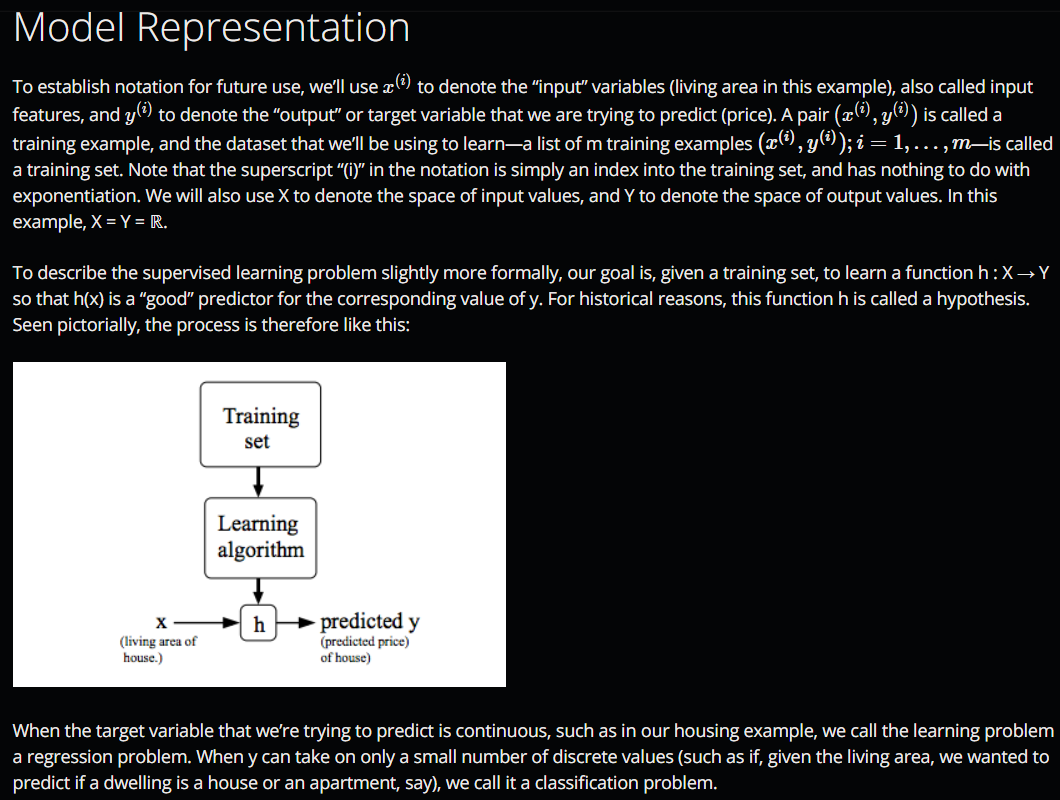
\*n. NLP natural language processing

supervised learning

n. Regression (target variable that we’re trying to predict is continuous)

n. Classification

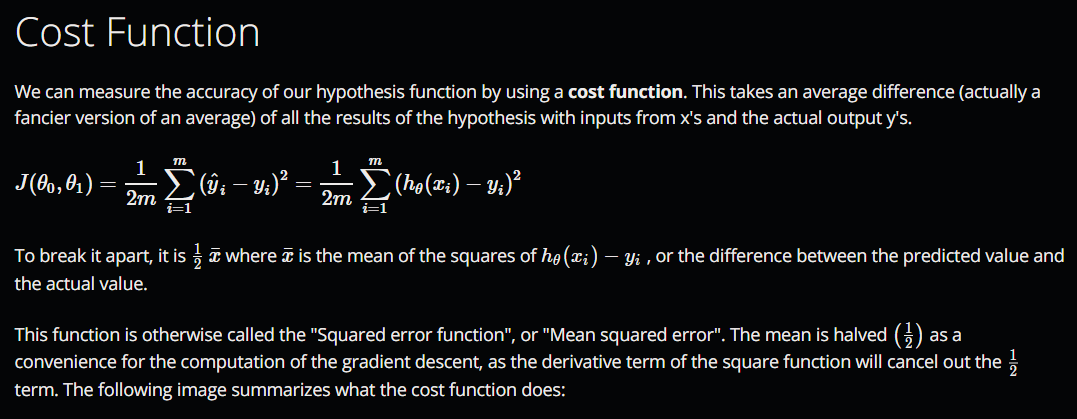
## Regression

通过date sets的学习得到hypothesis（假设），也即一个函数

\*n. superscript, subscript 上下标 dwelling 住址住宅， pictorial图像的，形象化的

问：为什么minimize这里要有一个1/2?

\*n. descent下降 derivative导数 recap扼要重述 low and behold你看，看吧

n. cost function用于衡量模型给出的预测值与实际值的偏差

n. gradient descent

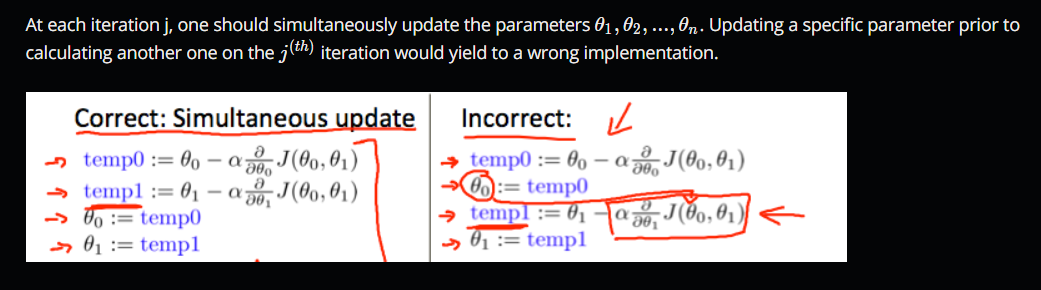
\*n. for the sake of brevity为了更加的简明扼要

\*n. Something that is **succinct** expresses facts or ideas clearly and in few words. 简明的 [表赞许]

\*n. If someone **asserts** a fact or belief, they state it firmly.

\*n. **Subtleties** are very small details or differences which are not obvious

词汇积累越来越多了呢

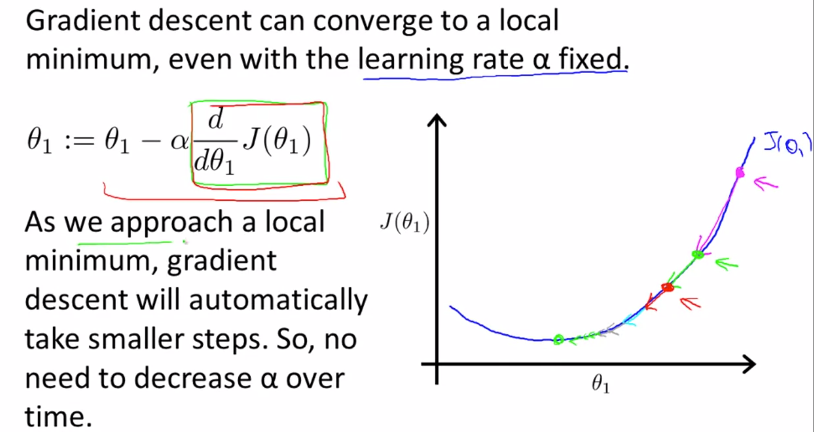
注意这个地方的simultaneous update！

\*n. To **depict** someone or something means to show or represent them in a work of art such as a drawing or painting. 描绘

\*n. slope斜率

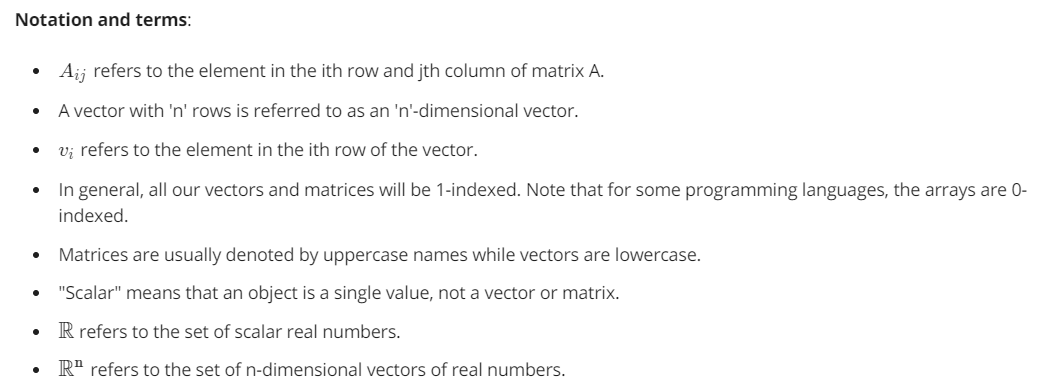
\*n. A **steep** slope rises at a very sharp angle and is difficult to go up.

\*n. To **convey** information or feelings means to cause them to be known or understood by someone

迭代的步进受两个因素影响，不仅受到learn rate的影响，而且导数项也随着接近local minimum而减小

n. Batch This method looks at every example in the entire training set on every step, and is called **batch gradient descent**.

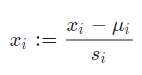
问：linear regression, nonlinear regression区别在哪里？

linear regression has only one global, and no other local, optima

关于矩阵以及向量

A non square matrix does not have an inverse matrix. Matrices that don't have an inverse are *singular* or *degenerate*.

Multiple features(variables)

1. Feature Scaling将feature都放在相近的区间可以加速收敛，取[-1,1]
2. Mean Normalization

Where *μ* is the **average** of all the values for feature (i) and s is the range of values (max - min), or s is the standard deviation.

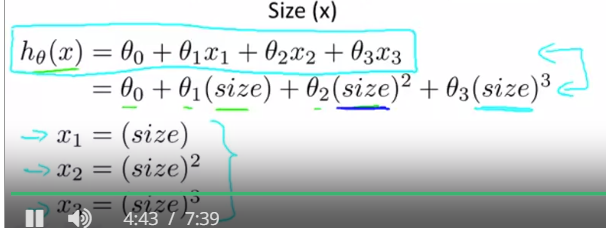
\*n. ellipse, oval都表示椭圆，注意区别

\*n. If you describe a sentence, idea, or system as **convoluted**, you mean that it is complicated and difficult to understand.

可以通过观察J再多次迭代后的变化来判断gradient descent是否正确工作，随之而来的是终止条件的设立

It has been proven that if learning rate **α is sufficiently small**, then J(θ) will decrease on every iteration.

关于learn rate选取的tradeoff

将高次幂的变量直接用新的变量替代

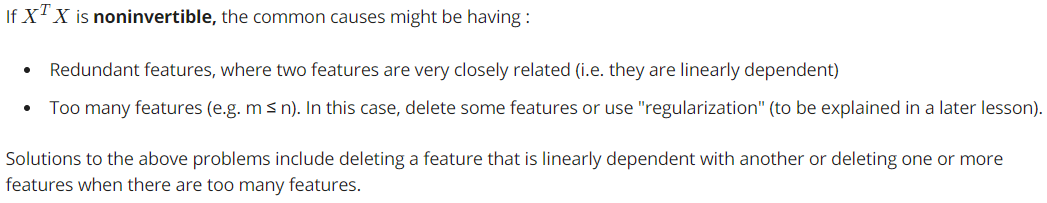
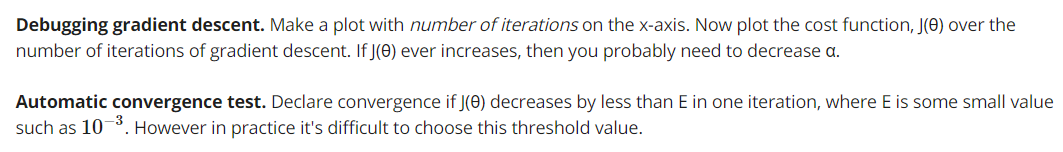
There is **no need** to do feature scaling with the normal equation.

n. Normal Equation

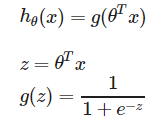
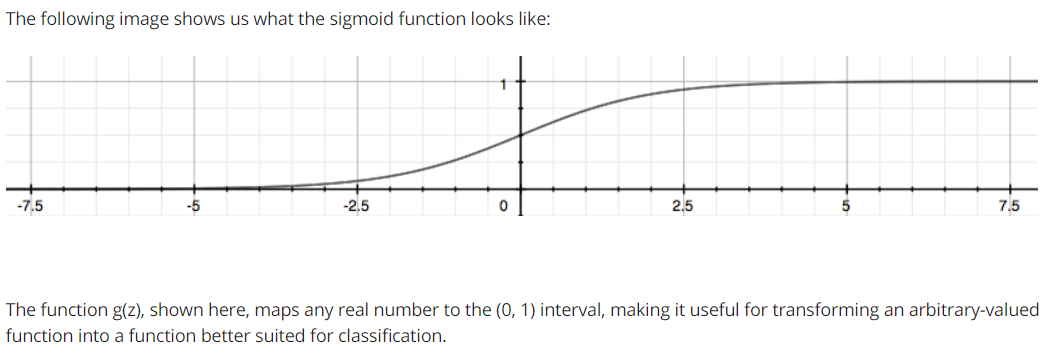
在Normal Equation和gradient descent两种方法中做选择

noninvertible cases

问：关于指令max(A, [], 1)不是很明白

问：A .\* 2和A .\* [2; 2; 2]为何结果相同

## Classification

-logistic function (sigmoid function) 满足了大于零小于一的要求，实际上相当于在原来的基础上增加了新的外函数，对函数值的范围做出了限定

-注意下式的含义，在给定参数为θ，输入为x时，y=1的可能性

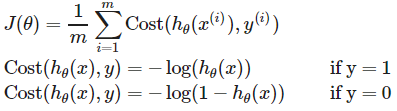
-在计算出可能性之后，取值规则是自己选的，比如当P大于0.5时取1，小于0.5时取0

\*n. magenta紫红色

\*n. If a person or group is penalized for something, they are made to suffer in some way because of it.

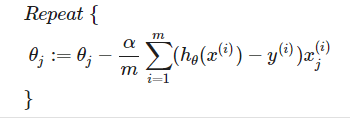
\*n. A wavy line has a series of regular curves along it. 波纹状的，波浪的

-Decision Boundary is the line that separates the area where y = 0 and where y =1，由参数theta所决定，与数据集无关。形状整个由z的表达式确定

-logistic regression cost function的详细分析，从本质上理解，利用了h(x)∈[0,1]的特性。可以将对数函数展开后理解

Note that writing the cost function in this way guarantees that J(θ) is **convex** for logistic regression.

-上述cost function可以用一个式子表示

- classification的gradient descent有着和linear regression相同的theta导数形式

-关于在octave中使用更加高级的优化算法，需要先给出一个特定返回值的costFunction()函数，然后使用fminunc()函数。

-Multiclass Classification，利用one-vs-rest方式，将每种class与剩余的所有classes做二分

## Overfitting

- It is usually caused by a complicated function that creates a lot of unnecessary curves and angles unrelated to the data.

-Bias, hypotheses have strong preconception

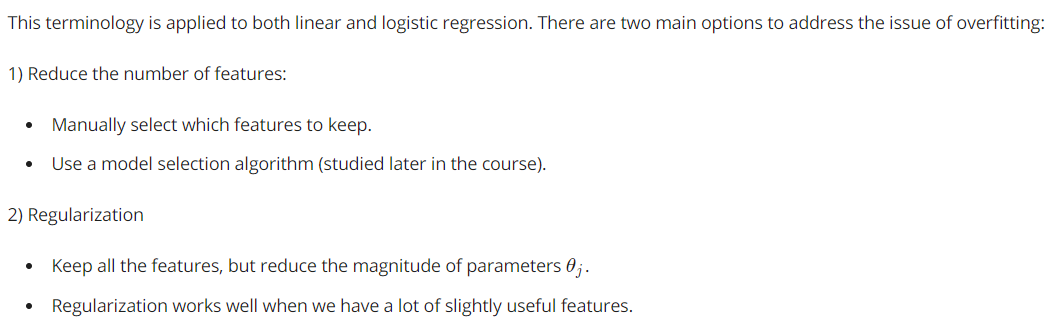
high bias, model is to simple to fit the data

just right,

high variance, doesn’t generalize well

-addressing overfitting

plot the hypotheses\*\*\*

reduce number of features

\*n. Your preconceptions about something are beliefs formed about it before you have enough information or experience.

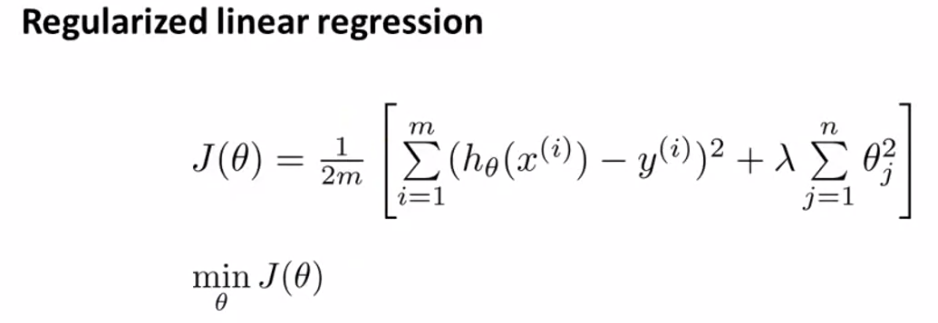
\*n. If a person or group is penalized for something, they are made to suffer in some way because of it.

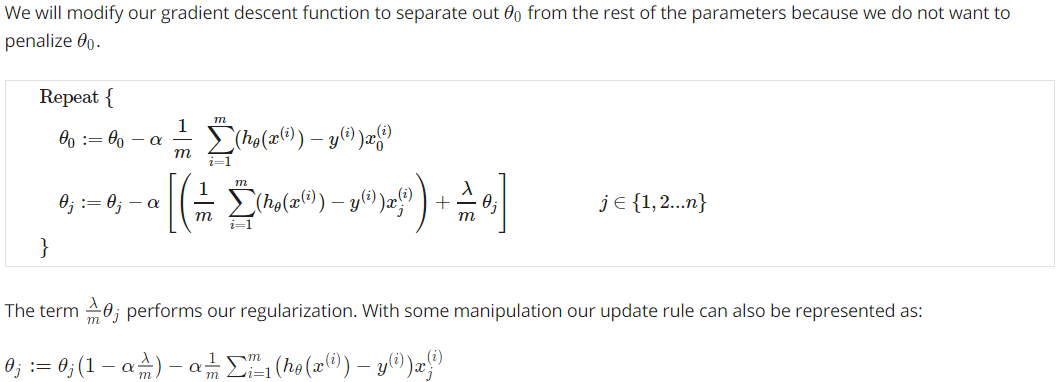
\*n. To be prone to something, usually something bad, means to have a tendency to be affected by it or to do it.

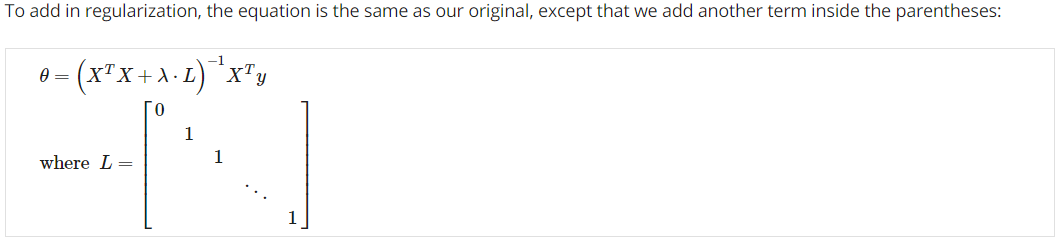
## Regularization

-regularization has two goals, fitting the training set well and keep the parameter small

-关于regularization中是否考虑theta0的问题

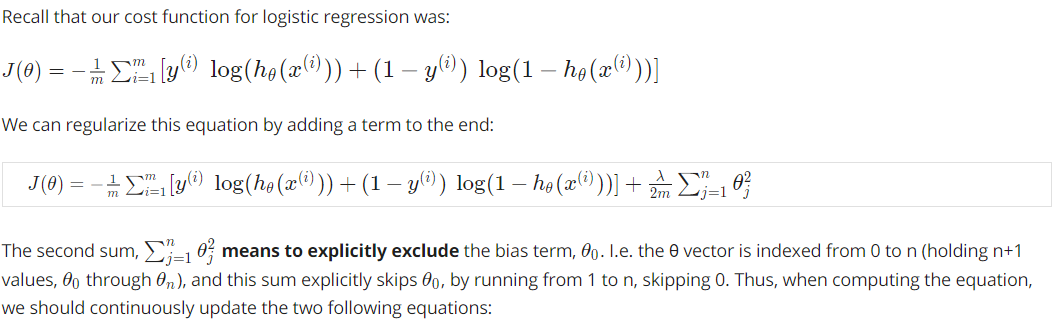
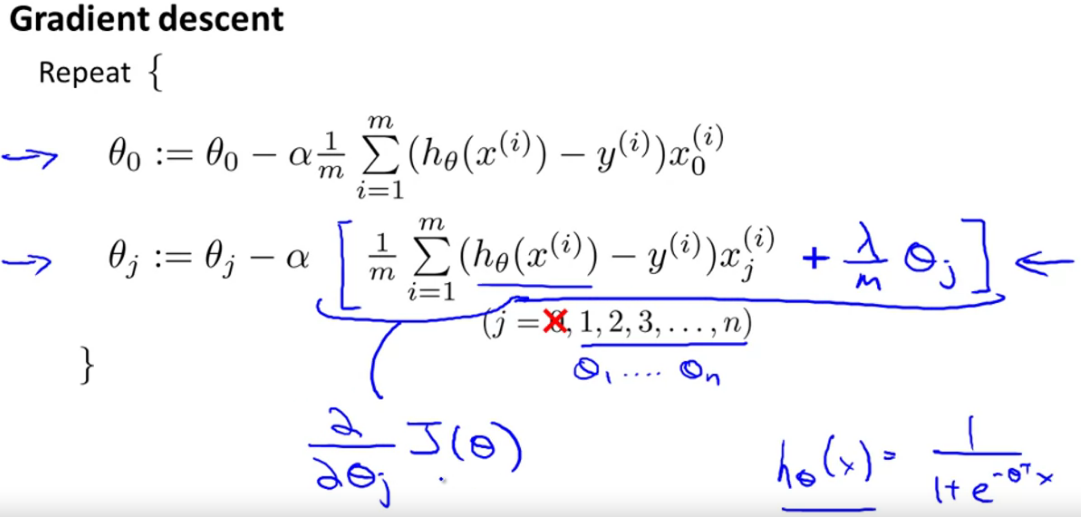
-在regularization下，新的cost function

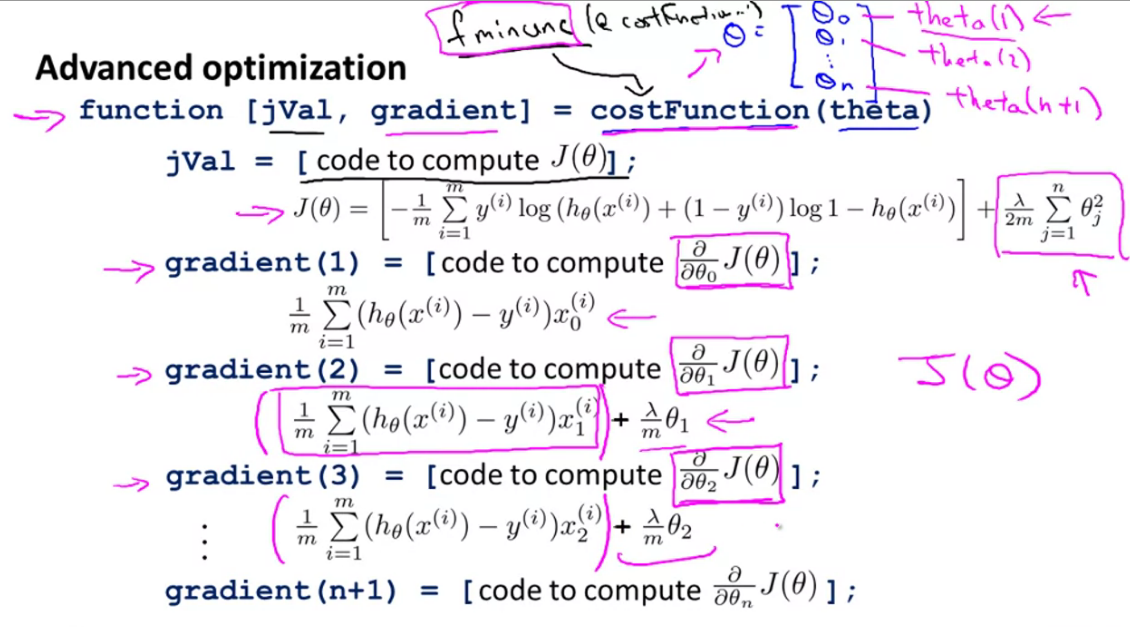
-在gradient descent中应用regularization

-在normal equation中应用regularization

the matrix L above has dimension as (n+1)×(n+1)

-using regularization can take care of non-inversible issue in normal equation

-在logistic regression中，cost function稍有不同，但是theta的更新仍然相同

-在advanced algorithms中利用regularization

-fminunc() is an optimization solver that finds the minimum of an unconstrained function.

-注意ex2种第99行代码，fminunc()的使用形式，以theta为变量的表达方式@(t)

-关于plot的绘图技巧还需要再好好看看

-说实话第二个练习中mapFeature部分的代码我没看懂

-关于contour()函数的使用，contour(Z,v)是以向量v中的数据来绘制矩阵Z的等高线。等高水平线的个数等于向量v的长度。如果想画一条单个等高线，使用contour(Z.[i i]).