第二周

Octave安装

多元线性回归

顾名思义，与单变量线性回归的区别是有多个特征值，即考虑多个影响因子。有以下符号标示：

表示第i个样本的第j个值。

表示第i行样本。

N表示一共有几个特征值。

多元线性回归预测函数：



可以简化为：

有向量：

简化为：，且为1。

多元线性回归的梯度下降

定义多元线性回归的预测函数：



定义梯度下降函数：

注意：同时更新，j=0,…,n

化简后梯度下降函数：

有=1，j=0,…,n

特征缩放

这是一种加速梯度下降的方法。假设当前有特征值，当特征值之间的范围大小相差很大时，则梯度下降会很慢，迟迟找不到最优点。特征缩放是指：把所有特征值用特征值除以某个值使得它们的范围落在-1<x<1或者-0.5<x<0.5（理想情况下），这个范围不是固定，视频中老师的经验是在-3<x<3都是可以接受的，但是对于-100<x<100或者-0.001<x<0.001的情况，则是不合适的。当实现合适的特征缩放时，梯度下降将会更高效，循环次数更少。这是针对特征值的范围实现加快梯度下降的方法。

均值归一化

这个是针对每一个特征值的处理，将特征值进行均值归一化。具体操作如下：



代表特征值的值，代表该特征的平均值，代表特征值的范围差（或者标准差）

如何判断梯度下降是否正确工作

判断有两种方法：图和自动收敛测试。

图：画出随迭代次数增多的曲线图。当学习速率太小，则会出现梯度下降很慢；当学习速率太大，则会出现梯度上升的现象，可能不会出现收敛。当学习速率适当时，会逐渐收敛于一个稳定的范围。

自动收敛测试：则是通过算法来判断，当减少至一个很小的固定值（预先指定，例如：）时，则说达到了收敛。

多项式回归

当使用线性回归不能更好的拟合样本数据，可以使预测函数变成二次方程、三次方程、平方根方程等，为了更好的拟合数据。例如：



但可以从不同的角度来看问题，可以增添新的特征值，例如：



这就由多项式回归（Polynomial Regression）变成了多元回归（Multiple Features）。此时由于特征的大小相差很大，所以特征缩放和均值归一化就比较重要。

标准方程

标准方程（Normal Equation）是另一种求最优化情况下的值的方法。定义如下：



X为设计矩阵，定义如下：



Y为样本数据正确结果的向量。

梯度下降和标准方程比较

梯度下降优点：无论样本规模大小或巨大，都可以解决问题，找到最优化情况下的值。梯度下降缺点：需要选择，值决定梯度下降的快慢，过大过小都不易；需要通过多次迭代才能得出结果，规模小时，相比比较慢。

标准方程优点：不需要选择；不需要迭代。标准方程缺点：求的复杂度为，n为样本规模，当样本规模多大，视频指出一般超过10000的情况（经验），就应该考虑梯度下降。

作业提交方法介绍

Octave基础使用方法

1. 加、减、乘、除（基本运算）
2. 与、或、非（逻辑运算）
3. 判断 == ~= （注释为%）
4. 调整命令行格式：PS1(‘>>’)
5. B=1;不会显示B，加分号。
6. Disp(sprint(‘2 decimals:%0.2f’,a));也是用来显示
7. 矩阵A=[1,2;3,4;5,6]，分号用来换下一行。
8. V=1:0.1:2；生成从1开始，以0.1为增量，到2的一个向量。
9. Ones(1,3)生成一个一行三列的所有元素都为1的矩阵。Zeros(1,3) 生成一个一行三列的所有元素都为0的矩阵。Rand(3,3) 生成一个三行三列的所有元素都为随机值（0-1）的矩阵。
10. 绘制直方图：hist(w)或者hist(w,50)
11. Eye(3)；创建一个3\*3的单元矩阵
12. Size(v)；输出该矩阵的维度。Size(v,1)输出该矩阵的行，size(v,2)输出该矩阵的列。
13. Length(v)；输出该矩阵的维度最大值。
14. Pwd用户显示当前octave的路径，用cd来改变当前的路径。通过load来加载问价 ，比如：load featureX.bat。who用来显示当前octave的变量，whos用来详细显示octave的变量及信息。Clear用来删除一个变量。
15. V=priceY(1:10);表示将priceY文件中的数据的前10行赋值到V中。Save hello.mat v;用来把v矩阵的数据保存到hello.mat中，load hello.mat可以再次加载v中。“：”表示该列或该行的所有数据。
16. 矩阵乘积和矩阵点乘，max方法，abs方法，log方法，exp方法，sum方法，prod方法，floor方法（向下取舍），ceil方法（向上取舍）。Eye(3)是3\*3的单位矩阵，flipud(3)是3\*3的反对角线单位矩阵。矩阵之间的max是比较两个矩阵元素之间的大小；矩阵内部的比较是取每列的最大值。
17. Magic(3)是构造一个3\*3的魔方矩阵，各行各列各对角线的和都是一个值。
18. Pinv是求矩阵的逆。
19. Plot画图命令。Xlabel标示x轴，ylabel标示y轴，title标示标题，legend标示图例。Print –dpng ‘myplot.png’保存图片。
20. Subplot(1,2,1)把控制框分成1\*2的格子，当前使用第一个。使用axis调整x/y轴的范围大小。使用clf清除数据。
21. While，for，if-else，break，continue使用方法。
22. 函数定义和调用：

注：求代价函数函数定义

向量化

对原来的梯度下降的进行向量化计算，如下：



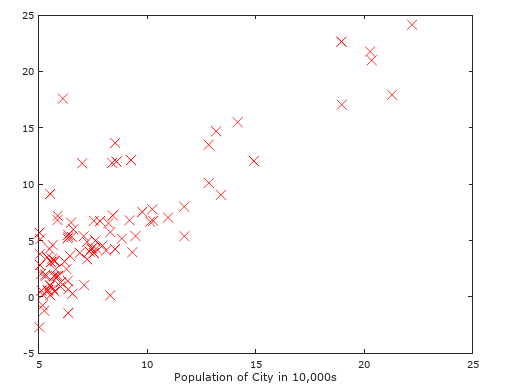
，，



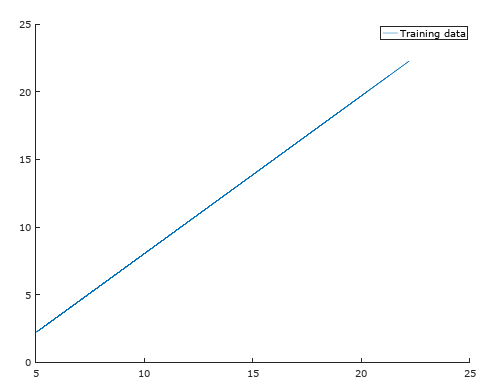
，，这样计算方便简洁，还满足同时更新要求。

作业：

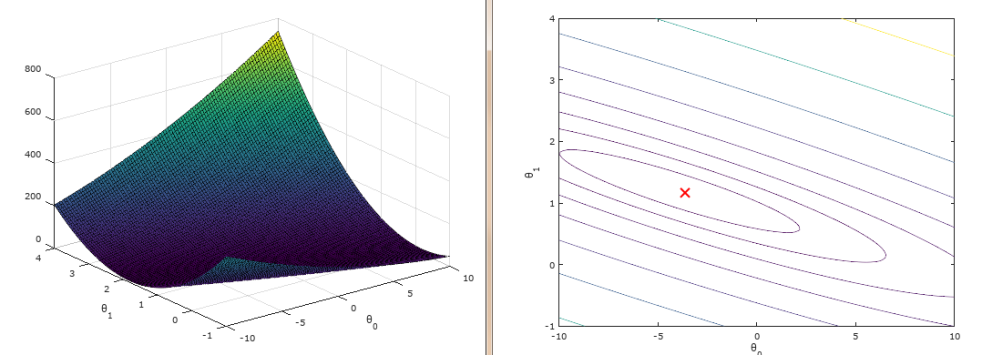
1.绘制X中数据的散点图



2．通过梯度下降算法，得到thera0=-3.6303，thera1=1.1664（1500次的迭代后，两个值趋于此值），如下图：



3.通过ex1.m函数画出的二维图和等值线图，显示出的最优值。



学习心得：

通过这周的学习，是对上周学习的拓展，上周是单变量线性回归，这周学习的是多变量线性回归和多项式回归。方法基本相同，梯度下降方法还能实用。并且又学习了标准方程的方法求解thera（不需要alpha和迭代），但是有规模的限制，当规模过大时，复杂度太大。这周还有最大的提升就是学习使用了Octave，了解了很多技巧和方法。通过作业，也感触颇深，感觉到理论和实践还是有一定差距的，不过还好有作业，这样可以加强巩固。通过作业，感觉到，很多计算都可以用矩阵来解决，方便简洁，并且Octave已经提供了矩阵的相关计算，所以这样是非常省事的，避免了很多循环计算。在做作业的过程中，发现自己在由传统计算向矩阵计算（也就是向量化的过程不熟练）转变的过程缺乏经验，不熟练。慢慢感觉知识难度变大，练习也有趣了，动手实践也多了。继续保持！