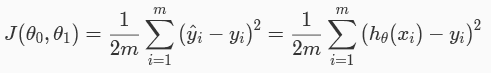
Machine Learning

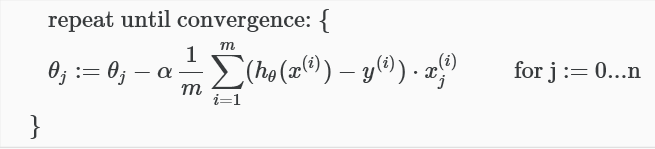
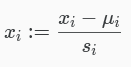
Introduction

1. 监督学习（supervised learning）：给出一堆已有数据和答案，让算法预测一组新数据的答案（一次决定）
2. 非监督学习（unsupervised learning）：只有一堆数据（找规律）；聚类
3. 强化学习（reinforcement learning）：奖惩机制；根据每次的结果自动进行优化（多次决定）

Linear Regression with One Variable

1. 成本函数（cost function）:
2. 梯度下降要注意“同时更新（simultaneously update）”的问题

Multivariate Linear Regression

1. 更新参数：
2. 特征缩放(feature scaling)：当各个特征的取值范围相差较大时，梯度下降算法收敛到最小值的速度很慢，可以通过特征缩放来加快梯度下降算法收敛的速度。一般应使特征取值范围为[-1,1]，但并非严格要求，相近即可。
3. 均值归一化（mean normalization）：
4. 

其中μi表示特征取值均值，si表示特征取值的极值或标准差

1. 通过作出 “J(θ)-迭代次数” 函数图像可以判断梯度下降算法是否正常工作。如果学习速度α选择得合适，每次迭代后J(θ)都会减小，如果J(θ)不随迭代次数单调递减，则应该适当缩小学习速度α。α也不宜过小，否则梯度下降算法的收敛速度会很慢。
2. 将已有的特征进行适当的组合得到新的特征有时可以达到很好的效果。
3. 正规方程：**θ = (XTX)-1XTY**

**X\*θ=Y（其中X表示特征取值矩阵，注意要加一项x0=1；Y表示相应特征取值的已知结果）。利用正规方程可以直接计算出θ0~θn的值。利用正规方程求解θ不必进行特征缩放。但是当n的值很大，即特征很多时，求解正规方程可能会很慢，此时可以考虑使用梯度下降方法。**

**8.**若X不可逆，有两种可能：采用的特征中有两个及以上特征具有相关性；特征太多而样本太少（如n=100, m=10）。解决方法分别是删除具有相关性的特征和删除一些特征以减少特征数量。

Octave/Matlab Tutorials

1. 指定小数位数：disp(sprintf(‘%0.2f’,a)； 使用format long 和 format short 命令可以调整默认显示小数位数
2. 等差行向量：v = 起始值:步长:结束值 如：v = 1:0.1:2
3. ones(m,n)：生成一个m行n列的矩阵，其中的元素全为1；类似地，zeros(m,n)可以生成一个m行n列的零矩阵；rand(m,n)可以随机生成一个m行n列的矩阵，元素数值均在0~1；randn(m,n)可以按照正态分布生成一个m 行n列的矩阵，其元素取值的服从正态分布，平均值为0
4. hist(m,n)：绘制柱状图，其中m为数据。例如一个一行a列的行向量，n为柱子的个数
5. eye(n)：生成一个n阶单位矩阵
6. size(A)：返回矩阵A的行数列数，返回的类型是一个一行两列的矩阵；size(A,1)和size(A,2)可以分别返回A矩阵的行数和列数；length(A)可以返回A矩阵的最大维数，即行数和列数中较大的那一个，但通常只对向量使用此函数，而不对一般矩阵使用
7. 使用cd “xx:\xxx\xxx”命令切换到指定文件夹后，可以用load example.dat命令或load(‘example.dat’)命令读取数据文件；who命令可以显示当前存储的全部变量；whos命令则不仅可以列出当前存储的全部变量，还可显示其详细信息；使用clear example命令可以删除example变量；使用save example.mat v命令可以把一个v变量存为一个名叫example.mat的文件，此命令会将数据以二进制格式存储，若想将数据存为普通的形式，可以使用命令save example.txt v –ascii
8. A(m,n)表示A矩阵第m行第n列的元素，A(m,:)表示A矩阵第m行的全部元素，A（:,n）表示A矩阵第n列的全部元素；A([m n],:)表示A矩阵的第m行和第n行的全部元素；A = [A,[a;b;c]]可以在A矩阵的右边增加一列[a;b;c]；A(:)可以将Ａ中的所有元素放入一个列向量；C = [A B]和C = [A;B]可以将A矩阵和B矩阵横向和竖向放在一起从而构成一个新矩阵
9. A .\* B表示A中的每一个元素与B矩阵对应位置上的元素相乘。点号通常表示位运算，所以类似地还有 A .^ 2 1 ./ A等操作
10. A’表示矩阵A的转置矩阵；magic(n) 幻方矩阵；find(a>b) 寻找指定条件元素；sum(A) 返回矢量各元素之和或矩阵每一列各元素之和；prod(A)返回矢量各元素之积或矩阵每一列各元素之积；floor(A)将A中的每一个元素缩小到小于等于该元素的最接近整数；ceil(A)将A中的每一个元素扩大到大于等于该元素的最接近整数；max(A,[],1)返回矩阵每一列的最大值，max(A,[],2)返回矩阵每一行的最大值；flipud(A)将A矩阵沿水平对称轴上下翻转
11. pinv(A)和inv(A)：求矩阵A的逆矩阵
12. plot()函数用来画二维图，有很多可选参数，详细信息可使用help plot查询；hold on命令可以保留当前已经画出的图形不被新的图形所覆盖；legend()函数用于在函数图像上添加图例；使用print –dpng ‘example.png’可以保存生成的图像，也可保存为其他格式的图片文件，详见help plot；subplot()函数可以将一张图片分为不同的格子，在每个格子里分别绘制不同的函数图像，便于相互对比；axis([a b c d])函数可以指定横坐标轴和纵坐标轴的范围（a,b表示横轴范围，c,d表示纵轴范围），还有其他参数可选
13. imagesc(A)可将A矩阵的数据以彩色方格图的形式呈现出来，颜色深浅代表值的大小，可以配合colorbar、colormap等命令使用

Multivariate Linear Regression

1. 少量数据：批梯度下降（batch gradient descent）
2. 大量数据：随机梯度下降（stochastic gradient descent）
3. 参数学习算法（parametric learning algorithm）：参数个数一定
4. 非参数学习算法（Non-parametric learning algorithm）：参数个数随着训练数据集合的增加而增加

（1）局部加权回归（locally weighted regression）