**重 庆 大 学**

**学 生 实 验 报 告**

**实验课程名称 数学实验**

**开 课 实验室 第一实验室DS1407**

**学 院 计算机学院 年级 2019 专业班 计卓02**

**学 生 姓 名 李燕琴 学号 20195633**

**开 课 时 间 2020 至 2021 学年第 二 学期**

|  |  |
| --- | --- |
| **总 成 绩** |  |
| **教师签名** |  |

**数 学 与 统 计 学 院 制**

**开课学院、实验室： 数统学院DS1407 实验时间 ： 2021年03月28日**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **课程**  **名称** | **数学实验** | **实验项目**  **名 称** | **拟合与多元线性回归** | **实验项目类型** | | | | |
| **验证** | **演示** | **综合** | **设计** | **其他** |
| **指导**  **教师** | **龚劬** | **成 绩** |  |  |  |  |  |  |
| 实验目的  5号宋体  基础实验1**拟合多项式**  问题重述  对文件中的x和y两个变量。编写脚本对其做第一、第二、第三、第四和第五次多项式拟合。仿照下图，用蓝色的点绘制x和y的散点图，然后在同一坐标系下，用不同颜色绘制多项式拟合曲线，并在合适的位置加入标注(xlabel,ylabel,legend)。要想获得拟合预测的误差估计，你需要用到polyfit（[P,S,MU] = polyfit(X,Y,N)）的第3输出MU：x的均值MU(1)和标准差MU(2)，详见help。x的均值和标准差可以对数据标准化（XHAT = (X-MU(1))/MU(2)），使其均值为0，方差为1。这个标准化变换改进了多项式和拟合算法的数值特性。还需要使用与之搭配的polyval（[Y,DELTA] = polyval(P,X,S,MU)）,该函数需要输入polyfit返回的S和 MU来计算拟合误差DELTA。  实验过程  load randomData.mat  plot(x,y,"k.");  legend("Data",'Position',[0.671309522208713 0.174365079989509 0.158214287315096 0.244047624497187]);  hold on  for i=1:5  [beta2,s,mu]=polyfit(x,y,i);  [y0,delta]=polyval(beta2,x,s,mu);  plot(x,y0,"DisplayName","Order"+i);  end  title("Polyfit to noisy data");  ylabel("Y")  xlabel("X")  实验结果及分析  实验结果如下图：可以看出，第五多项式的拟合效果最好，其次是第四多项式，之后依次递减，符合实验预期。  D:\2021study\1.课程学习\04-数学实验\hw_Ex\实验6-拟合与线性回归练习\n项式拟合结果图.png  图表 1 n项式拟合结果，n=1,2,3,4,5  基础实验2**经济增长模型**  问题重述  增加生产、发展经济所依靠的主要因素有增加投资、增加劳动力以及技术革新等，在研究国民经济产值与这些因素的数量关系时，由于技术水平不像资金、劳动力那样容易定量化，作为初步的模型，可认为技术水平不变，只讨论产值和资金、劳动力之间的关系。在科学技术发展不快时，如资本主义经济发展的前期，这种模型是有意义的。  用*Q，K，L*分别表示产值、资金、劳动力，要寻求的数量关系*Q(K,L)*。经过简化假设与分析，在经济学中，推导出一个著名的Cobb-Douglas生产函数：  *Q(K,L) = aKαLβ*， 0<α,β<1 （\*）  式中α,β*，a*要由经济统计数据确定。现有美国马萨诸塞州1900—1926年上述三个经济指数的统计数据，根据给定数据，试用数据拟合的方法，求出式（\*）中的参数α*,*β*，a*。  实验过程  clear,clc  % 把题目中所给的数据存入economicData中，（即持久化处理），在load进工作区  load economicData.mat  min\_resnorm = 1e9;  a = zeros(3,1);  % 通过循环，减少初始值选取的影响，确定最优值  for i =1:20  a0 = -5+10\*rand([3,1]);  [resa, resnorm, residual] = lsqcurvefit(@economy,a0,[K L],Q);  if resnorm<min\_resnorm  a = resa;  min\_resnorm = resnorm;  end  end  resQ = economy(a,[K L]);  a  % 求解标准差  err = sum((resQ-Q).^2)/size(Q,1)  function Q = economy(a,x)  %ECONOMY 题目给定的生产函数  K=x(:,1); L=x(:,2);  Q = a(1).\*K.^a(2).\*L.^a(3);  end  实验结果及分析  根据多元非线性拟合lsqcurvefit，求得参数，求解标准差验证结果的准确性，发现，err较小，系数拟合结果准确度高。  基础实验3 多元线性回归  问题重述  1）根据给定数据，对二者建立拟合回归模型，分别用以下方法构建模型：  方法一：最小化误差的平方和，即最小二乘法，求以下函数的最小值对应的最优解 即可。    可使用fminsearch进行求解，使用零作为初始值，绘制具有数据点的拟合线（具有beta中的系数）进行验证。  方法二：由于只有一个独立变量，也可以用**polyfit**对数据进行拟合。调用**polyfit**来求X和Y的拟合直线，并验证返回的系数是否与**fminsearch**中的系数匹配（因为**polyfit**也是基于最小二乘拟合）  方法三：因为最小二乘拟合有一个解析解，称为正态方程（可以通过求导并令其等于零得到）：    验证该公式是否能得到相同的系数集。  方法四：事实上，因为我们有一个超定的线性方程组，所以我们可以用左除”\”来解。验证你的答案。  方法五：用**regress**来估计参数，给出参数的区间估计，线性模型是否有效?  2）上述方法，拟合曲线会因为一些离群值而不能很好地拟合大多数数据，实际上，用**绝对偏差误差代替平方和误差**，可以显著提高模型对异常值的鲁棒性。使用该方法得到数据，将数据、最小二乘曲线（蓝色）和最小绝对偏差曲线（红色）绘制在一起。  3）探索目标函数使用其他p范数（公式如下）时会发生什么：    实验过程  %% 多元线性回归  load regression.mat  plot(X,Y,".")  hold on  X1 = ones(size(X,1),1);  xAug = [X X1];  beta0 = zeros(size(xAug,2),1);  %% =========方法一：平方和误差求最小值=========  beta1 = fminsearch(@(beta0)squaredCost(beta0,xAug,Y),beta0)  resY1 = xAug\*beta1;  plot(X,resY1);  hold off  %% =========方法二：一个独立变量，可使用polyfit进行拟合=========  beta2 = polyfit(X,Y,1)  %% =========方法三：最小二乘拟合的解析解（正态方程）=========  beta3 = (xAug'\*xAug)\xAug'\*Y  %% =========方法四：超定线性方程组，左除求解=========  % xAug \* beta = Y; 则 beta = xAug \ Y  beta4 = xAug \ Y  %% =========方法五：regress求解=========  beta5 = regress(Y,xAug)  %% =========绝对偏差误差代替平方和误差，以减少异常值============  beta6 = fminsearch(@(beta0)absoluteCost(beta0,xAug,Y),beta0)  resY6 = xAug\*beta6;  plot(X,Y,"k.",X,resY1,"b",X,resY6,"r");  legend("原始点","平方和误差","绝对偏差误差")  %% =========选用不同范数p实验==========  plot(X,Y,'k.');  hold on  for i=1:5  beta = fminsearch(@(beta0)Cost(beta0,xAug,Y,i),beta0);  resY = xAug\*beta;  plot(X,resY);  end  legend('原始点','p=1','p=2','p=3','p=4','p=5')  %% =========相关函数==============  function cost=squaredCost(beta,dataAug,labels)  cost = sum((dataAug\*beta-labels).^2);  end  function cost=absoluteCost(beta,dataAug,labels)  cost = sum(abs(dataAug\*beta-labels));  end  function cost=Cost(beta,dataAug,labels,p)  cost = sum(abs(dataAug\*beta-labels).^p);  end  实验结果及分析  1、使用上述五种方法，得到结果均相同，其中;  D:\2021study\1.课程学习\04-数学实验\hw_Ex\实验6-拟合与线性回归练习\untitled2.png  图表 2 最小二乘方法拟合  2、使用绝对偏差代替平方和误差，发现前者会很好地拟合大部分数据，受异常值的干扰较小，比较结果如下图：  D:\2021study\1.课程学习\04-数学实验\hw_Ex\实验6-拟合与线性回归练习\最小二乘法与绝对偏差法的拟合效果比较.png  图表 3 最小二乘法与绝对偏差法的拟合效果比较  3、通过比较误差向量的不同范数p下的损失值求最小，以得到合适的系数值，发现p越大，拟合效果受异常值影响越大。  D:\2021study\1.课程学习\04-数学实验\hw_Ex\实验6-拟合与线性回归练习\误差向量不同范数p下的拟合效果.png  图表 4 误差向量不同范数p下的拟合效果  原理总结分析：当p越大时，异常值p级数后，对误差的贡献就越大，导致拟合效果受异常值的影响变大。    总结与体会  **实验1**  注意，polyfit求得的参数p可以直接带入polyval进行求解。  **实验2**  Lsqcurvefit受初始值的选取的影响比较大，可以基于蒙特卡洛模拟的方法，采用for循环，选取误差最小的值，作为最终结果。  **实验3**  计算矩阵乘法时，维度计算需要仔细，如以下公式的推导    教师签名  年 月 日 | | | | | | | | |