山东大学 计算机科学与技术 学院

机器学习与模式识别 课程实验报告

|  |  |  |  |
| --- | --- | --- | --- |
| 学号：201600301304 | 姓名：贾乘兴 | | 班级： 人工智能16 |
| 实验题目：线性回归 | | | |
| 实验学时：2小时 | | 实验日期： 2018/9/28 | |
| 实验目的：通过梯度下降的方法不断训练线性回归中的参数，并绘制误差的分布来观察参数下降的方向，记录梯度下降时参数变换，最后利用训练参数对3.5与7岁身高进行预测 | | | |
| 硬件环境：mac os | | | |
| 软件环境：matlab 2017b | | | |
| 实验步骤与内容：  一．导入数据并显示  1.所给数据中，ex1x.dat为x数据，ex1y.dat为y标签，导入并绘制点的分布，代码如下：  %% show the data  x=load('ex1x.dat');  y=load('ex1y.dat');  figure  plot(x,y,'o');  xlabel('age in years');  ylabel('height in meters');  2.得到数据的规模，并进行处理，代码如下：  m=length(y);  X=[ones(m,1),x];   1. 训练定义   1.目标函数定义：本实验目的为线性回归，故目标函数为y=wx+b，对函数进行简化，可表示为y=X\*theta,X为m\*2维矩阵，theta为2\*1列向量  2.损失函数定义：采用最小二乘的定义，损失函数定义如下    其中h（theta）为假设，本实验为线性函数    3.训练目标为最小化损失函数，通过得到最优的theta从而得到最终的目标函数，本实验采用的是梯度下降的方法，J(theta)对theta求导可得    theta的更新公式为：    其中alpha为学习率，本实验采用的学习率为0.07，最终得到theta的更新为    本实验设置训练迭代次数为1500，（0，1）内设置随机的初始值   1. 实验过程   1.绘制损失函数的图像，代码如下：  %% loss=1/2\*sum(x\*theta-y)^2  Jtheta=zeros(100,100);  theta0=linspace(-3,3,100);  theta1=linspace(-1,1,100);    for p=1:length(theta0)  for q=1:length(theta1)  t=[theta0(p);theta1(q)];  Jtheta(p,q)=(X\*t-y)'\*(X\*t-y)/(2\*m);  end  end  figure  surf(theta0,theta1,Jtheta')  xlabel('\theta\_0');  ylabel('\theta\_1');  figure  contour(theta0,theta1,Jtheta',1000)  xlabel('\theta\_0');  ylabel('\theta\_1');  绘制图像结果如下，可大致观察到最低点的大致位置，其中  theta = [theta0;theta1]  ml/e1/result0.jpg  绘制等高线图如下：  result000.jpg  2.训练过程更新theta代码如下：  %% gradient descent  theta=zeros(2,1);  num=1500;  alpha=0.07;  for k=1:num  theta=theta-alpha\*sum((X\*theta-y).\*X)'/(m);  end  hold on;  plot(X(:,2),X\*theta,'-o');  3.本实验为线性回归，由损失函数定义与目标函数可知，我们的目的为找到最终的theta，使得    故由最小二乘可知，我们可以直接得到最优的theta，即：    其中X+为X的伪逆    所得到theta为最优的  将实验原始数据，梯度下降训练结果，最小二乘的结果绘制成图像进行对比，代码入下：  %% show the data  ......(代码如上)    %% show the train(y=Xw,w=inv(X'X)\*X')  theta=pinv(X)\*y;  hold on;  plot(X(:,2),X\*theta,'-x');    %% gradient descent  ......(代码如上)  %% plot legend  legend('training data','linear regression','GD')  得到的图像为  ml/e1/result.jpg  通过对比可知梯度下降的收敛效果很好，与精确的最优解基本一致   1. 实验结果 2. 实验结果保留四位小数，所得梯度下降和线性回归结果相等，theta0=0.7502 theta1=0.0639,最小二乘误差为9.8707e-04 3. 实验结果保留四位小数在初始值为[0,0]，学习率0.07下，theta第一次迭代的数值为[0.0745;0.3800] 4. 实验结果保留四位小数，对年龄为3.5与7岁进行预测，得到身高分别为0.9737m与1.1973m | | | |
| 结论分析与体会：通过本次实验，对梯度下降方法有了初步的了解，该方法在函数不复杂的情况下可以较为有效的收敛于全局最优，但当函数复杂的情况下，结果对参数依赖较大，同时我们可以使用一些数值优化方法对梯度下降进行调整，比如加入物理中动量的模型 | | | |

附录：程序源代码

ex1.m

clear,clc;

%% show the data

x=load('ex1x.dat');

y=load('ex1y.dat');

figure

plot(x,y,'o');

xlabel('age in years');

ylabel('height in meters');

m=length(y);

X=[ones(m,1),x];

%% show the train(y=Xw,w=inv(X'X)\*X')

theta=pinv(X)\*y

(X\*theta-y)'\*(X\*theta-y)/(2\*m)

hold on;

plot(X(:,2),X\*theta,'-x');

%% gradient descent

theta=zeros(2,1);

alpha=0.07;

for k=1:1500

theta=theta-alpha\*sum((X\*theta-y).\*X)'/(m);

if k==1

theta

end

end

hold on;

plot(X(:,2),X\*theta,'-o');

legend('training data','linear regression','GD')

%% loss=1/2\*sum(x\*theta-y)^2

Jtheta=zeros(100,100);

theta0=linspace(-3,3,100);

theta1=linspace(-1,1,100);

for p=1:length(theta0)

for q=1:length(theta1)

t=[theta0(p);theta1(q)];

Jtheta(p,q)=(X\*t-y)'\*(X\*t-y)/(2\*m);

end

end

figure

surf(theta0,theta1,Jtheta')

xlabel('\theta\_0');

ylabel('\theta\_1');

figure

contour(theta0,theta1,Jtheta',1000)

xlabel('\theta\_0');

ylabel('\theta\_1');

%% pre

y1=[1,3.5]\*theta

y2=[1,7.0]\*theta