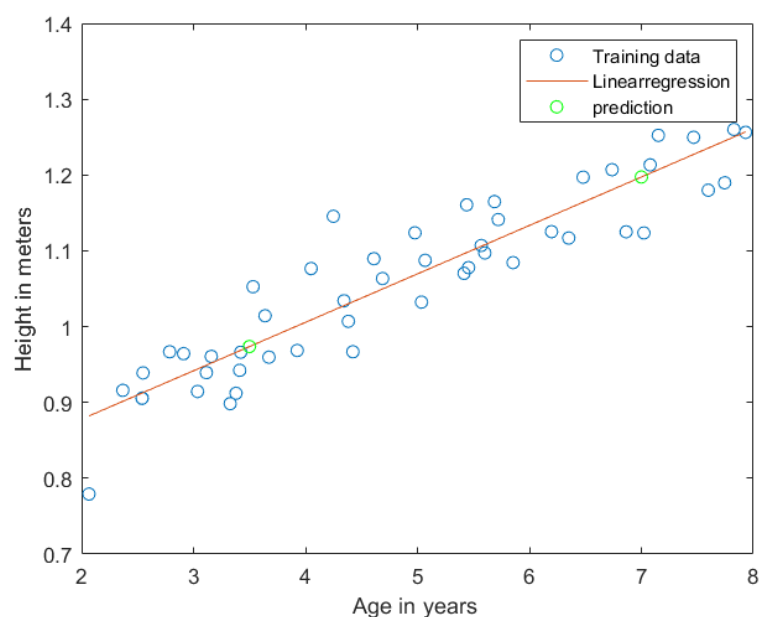


山东大学计算机科学与技术学院

机器学习与模式识别课程实验报告

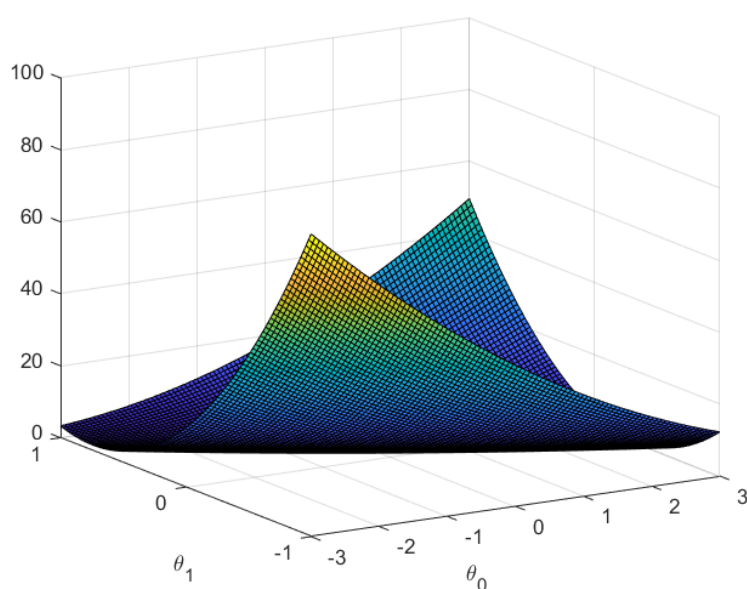
学号：201600181073	姓名：唐超	班级：16 人工智能
实验题目：Linear Regression		
实验学时：2	实验日期：2018. 9. 28	
<p>实验目的：</p> <p>已给出 50 组 2 到 8 岁男孩的年龄和相应的身高数据文件，用 MATLAB 建立一个年龄和身高的线性回归模型。</p>		
<p>硬件环境：</p> <p>DELL 台式机</p>		
<p>软件环境：</p> <p>MATLAB2016a</p>		
<p>实验步骤与内容：</p> <ol style="list-style-type: none">1. 下载、解压实验所需的数据集，并加载到 MATLAB 中，其中年龄数据作为 50 维列向量 x，身高数据作为 50 维列向量 y。2. 以身高为纵坐标，以年龄为横坐标，画出散点图，观察其分布特征；3. 为了使回归方程中的常数项能统一用矩阵的乘法描述，在 x 的第一列前加上一列 1，成为 50×2 的矩阵；4. 建立线性回归模型： $h_{\theta}(x) = \theta^T x = \sum_{i=0}^n \theta_i x_i \tag{1}$ <p>其中 $n=1$, x 为矩阵 x 的某一行，θ 是我们要求的线性回归方程的系数组成的向量。</p> <p>根据损失函数 $J(\theta)$ 对 θ 的导数为零，有 θ 梯度下降的迭代公式如下：</p> $\theta_j := \theta_j - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)} - y^{(i)})x_j^{(i)} \tag{2}$ <p>初始化 $\theta = (0,0)^T$，$\alpha = 0.07$，经过第一次迭代，$\theta=(0.0745,0.3800)^T$，1500 次迭代后 θ 几乎稳定在 $\theta = (0.7502,0.0639)$，即为所求。在原来的散点图上画出 $h_{\theta}(x)$：</p>		



5. 通过已得到的回归方程预测 3.5 岁和 7 岁男孩的身高，代入回归方程得预测值分别为 0.9737m 和 1.1973m，在上图中以绿色散点标出。
6. 绘制并理解线性回归的损失函数 $J(\theta)$

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 \quad (3)$$

在 MATLAB 中，在横坐标范围为 $[-3, 3]$ ，纵坐标范围为 $[-1, 1]$ 的区域内，建立 100×100 的网格，绘制出 $J(\theta)$ 如下：



结论分析与体会：

本实验首先根据两组数据的相关关系建立一元线性回归方程 $h_{\theta}(x) = \theta^T x = \sum_{i=0}^n \theta_i x_i$ ，在损失函数 $J(\theta)$ 最小的约束条件下，求解 θ ，通过梯度下降的方法，经过约 1500 次迭代， θ 几乎不再变化，将此时的 θ 代回 $h_{\theta}(x)$ ，在原两组数据的散点图中画出，发现大部分点均匀地分布于直线 $h_{\theta}(x)$ 两侧，说明效果良好。

通过这次实验，我加深了对线性回归的理解，对 MATLAB 的使用也更加地熟练了。

附录：程序源代码

```
x = load('ex1x.dat');  
  
y = load('ex1y.dat');  
  
  
figure % open a new figure window  
  
plot (x , y , ' o ' ) ;  
  
ylabel ( ' Height in meters ' );  
  
xlabel ( ' Age in years ' );  
  
  
  
m = length ( y ) ;  
  
x = [ones(m,1),x] ; % Add a column of ones to x  
  
alpha = 0.07;  
  
theta = [0 0];  
  
  
  
for i = 1:1500  
  
    theta = theta-(alpha/m)*(theta*x'-y')*x;  
  
end  
  
  
  
hold on  
  
plot ( x ( : , 2 ) , x*theta' , '-')
```

```

% prediction

h1 = theta*[1 3.5]';

h2 = theta*[1 7]';

hold on

scatter([3.5 7],[h1 h2],'g')

legend('Training data' , 'Linearregression','prediction');

```

```

%cost function

J_vals = zeros(100,100);

theta0_vals = linspace(-3,3,100);

theta1_vals = linspace(-1,1,100);

for i = 1:length(theta0_vals)

    for j= 1:length(theta1_vals)

        t = [theta0_vals(i),theta1_vals(j)];

        J_vals(i,j) = (1/m)*(t*x'-y)*(t*x'-y)';

    end

end

J_vals = J_vals'; %surf(Z)创建曲面，并将 Z 中元素的列索引和行索引分别用作 x 坐标和 y 坐标。因此需要转置。

figure;

surf(theta0_vals,theta1_vals,J_vals);

xlabel('\theta_0');

ylabel('\theta_1');

```