**线性回归模型实验指导书**

一、实验目的

1. 熟悉如何通过Matlab实现机器学习算法；
2. 了解线性回归模型相关知识。

二、实验环境

系统：windows7

软件：Matlab

三、实验内容

1. 一元线性回归

数据集ex1data1.txt,有97个样本数据，每个样本有两列：人口和经济效益，通过实验，熟悉Matlab的一些基本函数和数据可视化方法。用最小二乘“参数估计”方法对模型进行求解。

2.多元线性回归

数据集ex1data2.txt中，有47个样本数据，每个数据有3列：房子大小，房间数目，房价。通过房子大小和房间数目作为样本特征，房价作为样本输出标签，通过线性回归模型来预测房价。请预测一个三间房，1650平方尺的房子的房价。

四、实验步骤

1. 一元线性回归：
2. 导入数据集1中的数据，作图表示数据集1中的数据图。

参考代码：

fprintf('Plotting Data ...\n')

data = load('ex1data1.txt');

X = data(:, 1);

y = data(:, 2);

m = length(y); % number of training examples

plot(X, y,'rx','MarkerSize', 10); % Plot the data

ylabel('Profit in $10,000s'); % Set the yaxis label

xlabel('Population of City in 10,000s'); % Set the xaxis label

fprintf('Program paused. Press enter to continue.\n');

pause;

1. 用最小二乘参数估计进行模型求解。

编程实现：

1. 根据教材P54页公式(3.7),(3.8)，求出w和b的最优解。
2. 根据公式(3.3)计算出y的预测值。
3. 作图画出原始数据和预测结果，进行对比（建议通过help plot，熟悉plot语句用法，比如横轴、纵轴名称，用不同的颜色和形状区分原始数据和预测结果等）

提示：使用Matlab中的语句：sum, 平方^2,注意矩阵中.\*和\*的区别。

1. 多元线性回归：
2. 根据教材P55页多元线性回归原理，用最小二乘法对参数进行估计，根据公式(3.11)求出最优解，学得模型。
3. 预测房价，比如：一个三间房，1650平方尺的房子的房价。

参考程序，其中有些部分需要同学补充完成。

data = load('ex1data2.txt');

X = data(:, 1:2);

y = data(:, 3);

m = length(y);

% Print out some data points

fprintf('First 10 examples from the dataset: \n');

fprintf(' x = [%.0f %.0f], y = %.0f \n', [X(1:10,:) y(1:10,:)]');

fprintf('Program paused. Press enter to continue.\n');

pause;

% Add intercept term to X

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

% Calculate the parameters from the normal equation

% 调用normalEqn函数，用正规方程解法计算参数，请同学们在normalEqn.m中完成代码

theta = normalEqn(X, y);

% Display normal equation's result

fprintf('Theta computed from the normal equations: \n');

fprintf(' %f \n', theta);

fprintf('\n');

% 请预测一个三间房，1650平方尺的房子的房价。

% ====================== 请同学们在下面输入代码 ======================

price = 0; % 请修改这条语句，得到预测结果。

% ============================================================

fprintf(['Predicted price of a 1650 sq-ft, 3 br house ' ...

'(using normal equations):\n $%f\n'], price);

normalEqn.m代码：

function [theta] = normalEqn(X, y)

%NORMALEQN Computes the closed-form solution to linear regression

% NORMALEQN(X,y) computes the closed-form solution to linear

% regression using the normal equations.

theta = zeros(size(X, 2), 1);

% === 下面语句请同学们完成，通过用matlab的pinv函数求解线性回归的闭式解=======

% Instructions: Complete the code to compute the closed form solution

% to linear regression and put the result in theta.

%

end