

# 实验一 线性回归

**(特别说明:**

**1、请按照实验分组完成实验，实验采取线上方式，需在11月16日前完成验收！**

**2、各小组需在验收之前将实验报告、程序等压缩为一个文件发给助教！（助教信息会在群里公布）**

**3、验收之后，各组组长将最终实验报告等发给各班课代表，由各班课代表在11月18日前统一提交给助教！**

**文件压缩包命名方式：计科\*班+组长+成员名单)**

本实验要求应用线性回归模型解决实际问题，通过实验加深对线性回归原理的理解。建议使用python 编程实现。

实验包括两部分：一元线性回归和多元线性回归。按百分制计算，第一部分80分，第二部分20分。（两部分实验均完成较好的同学会获得更高分数）。

## 1、一元线性回归

题目：

应用一元线性回归预测移动餐车的利润。假设你是一家餐饮连锁店的CEO，考虑在不同的城市开辟新店。该餐饮店已在许多城市拥有移动餐车，现有各个城市移动餐车的利润和城市人口的数据。这些数据将帮助你选择在哪个城市进行新店扩张。请按要求完成实验。

数据集：

文件ex1data1.txt 为该实验部分的数据集，第一列表示城市人口（单位为万人），第二列表示该城市的移动餐车的利润（单位为万美元，若利润为负值，表示损失）。

步骤与要求：

- 1) 在开始任务之前，进行数据的可视化对于了解数据特征是很有帮助的。请你导入数据并以人口为横坐标，利润为纵坐标画出散点图并观察数据分布特征。（建议：用python 的matplotlib）
- 2) 将线性回归参数初始化为0，然后计算代价函数（cost function）并求出初始值。
- 3) 使用线性回归的方法对数据集进行拟合，并用梯度下降法求解线性回归参数。（eg: 迭代次数=1500,  $\alpha=0.01$ ）
- 4) 画出数据的拟合图形。
- 5) 预测人口数量为35000 和70000 时，利润为多少。

## 2、多元线性回归

题目：

应用多元线性回归预测房价。假设你打算出售你的房子，你想知道房子的市场价应该设为多少比较合适。一种方法就是收集最近的房屋销售信息并设计一个房屋价格模型。请按要求完成实验。

数据集：

文件exldata2.txt 为该实验部分的数据集，第一列表示房屋的面积（平方英尺），第二列表示房间数目，第三列表示房屋价格。

步骤与要求：

- 1) 导入数据，通过观察，容易发现房屋面积的大小约是房间数量的1000 倍。当特征数量级不同时，对特征缩放能够使梯度下降更快地收敛。请对这两个特征进行归一化处理。
- 2) 使用梯度下降法求解线性回归参数。尝试使用不同的alpha（学习率）进行实验，找到一个合适的alpha 使算法快速收敛。思考alpha 的大小对于算法性能的影响。
- 3) 使用你认为最佳的alpha 运行梯度下降法求出线性回归参数，然后预测房屋面积为 1650 平方英尺，房间数量为 3 时，房屋的价格。