# 机器学习-回归（课后作业）

## 实现原理

**最小二乘法：**

最小二乘法（又称最小平方法）是一种数学优化技术。它通过最小化误差的平方和寻找数据的最佳函数匹配。利用最小二乘法可以简便地求得未知的数据，并使得这些求得的数据与实际数据之间误差的平方和为最小。最小二乘法还可用于曲线拟合。其他一些优化问题也可通过最小化能量或最大化熵用最小二乘法来表达。

在我们研究两个变量（x,y）之间的相互关系时，通常可以得到一系列成对的数据（x1,y1.x2,y2... xm,ym）；将这些数据描绘在x -y直角坐标系中，若发现这些点在一条直线附近，可以令这条直线方程如（式1-1）。

其中：a0、a1 是任意实数

为建立这直线方程就要确定a0和a1，应用《最小二乘法原理》，将实测值Yi与利用计算值Yj（Yj=a0+a1Xi）（式1-1）的离差（Yi-Yj）的平方和 最小为“优化判据”。

令：φ = （式1-2)

把（式1-1）代入（式1-2）中得：

φ = （式1-3)

当 最小时，可用函数 φ 对a0、a1求偏导数，令这两个偏导数等于零。

（式1-4)

（式1-5)

亦即：

 （式1-6)

 （式1-7)

得到的两个关于a0、 a1为未知数的两个方程组，解这两个方程组得出：

a0 = （∑Yi) / n - a1（∑Xi) / n （式1-8)

a1 = [n∑(Xi Yi) - （∑Xi ∑Yi)] / (n∑Xi^2 -∑Xi∑Xi)（式1-9)

这时把a0、a1代入（式1-1）中， 此时的(式1-1）就是我们回归的一元线性方程即：数学模型。

在回归过程中，回归的关联式不可能全部通过每个回归数据点（x1,y1. x2,y2...xm,ym），为了判断关联式的好坏，可借助相关系数“R”，统计量“F”，剩余标准偏差“S”进行判断；“R”越趋近于 1 越好；“F”的绝对值越大越好；“S”越趋近于 0 越好。

R = [∑XiYi - m （∑Xi / m）（∑Yi / m)]/ SQR{[∑Xi2 - m （∑Xi / m)2][∑Yi2 - m （∑Yi / m)2]} （式1-10)

在（式1-10）中，m为样本容量，即实验次数；Xi、Yi分别为任意一组实验数据X、Y的数值。

**梯度下降算法：**

目标：给一个初始值，然后逐步的迭代改变的值，是代价损失函数逐次变小，使每次都往梯度下降的方向改变，其实就是每次迭代都需要计算所有样本的残差并加权。

下降速度：

求偏导（一个样本）：







即：

求偏导（多个样本），算法为：

Repeat until convergence{

 (for every j)

}

## 程序结构

**最小二乘法：**

①multivariatel\_liner\_fitting在完成拟合曲线参数计算前，对相应变量进行计算；

②calculate\_parameter利用高斯消元法，求得拟合曲线的参数；

③calculate得到第②步的参数后，对原数据进行回归预测，得到预测值列表；

④取得了原数据以及第③步的预测值后，完成函数和散点图的绘制，如图1。

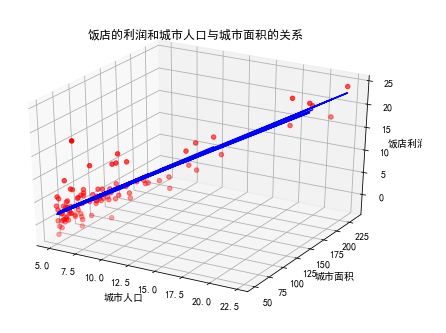


图 1

**梯度下降法：**

①load\_data加载数据，使变量以[城市人口,城市面积]、[饭店利润]的形式进行计算；

②featureNormalize进行特征缩放，得到均值和sigma；

③computeCost计算损失；

④gradientDescent进行梯度下降计算，得到参数集theta以及迭代过程中的全部loss；

⑤第④步得到的参数集进行数据预测，得到回归线上的点

⑥将第③步的loss进行画图，得到loss的迭代变化如图2，将第⑤步的预测值和原数据进行散点图和回归结果进行作图，如图3。

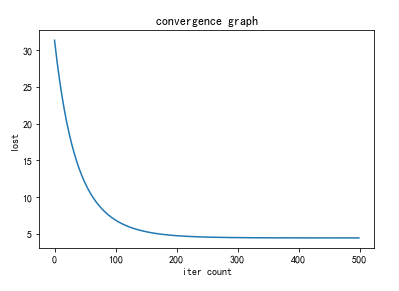


图 2

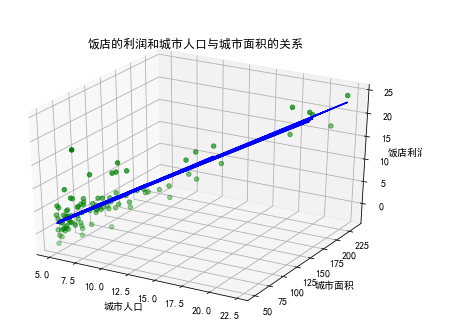


图 3

**对比：**

**相同点：**

1： 从结果上看用最小二乘法的训练模型的效果与梯度下降法模型训练的效果基本一致

2：两种方法都是在给定已知数据（independent & dependent variables）的前提下对dependent variables算出出一个一般性的估值函数。然后对给定新数据的dependent variables进行估算。

3：都是在已知数据的框架内，使得估算值与实际值的总平方差尽量更小

**不同点：**

最小二乘法是直接对△求导找出全局最小，是非迭代法。而梯度下降法是一种迭代法，先给定一个β，然后向△下降最快的方向调整β，在若干次迭代之后找到局部最小。

**参数改变:**

参数改变也会引起拟合效果的差异，比如学习率越低，那么需要的迭代次数就越高，否则造成拟合的回归线不精确，如图4和图5分别在学习率为0.001，0.01，迭代步数为1000的情况下进行的数据拟合，可以看出图5比图4更加准确。

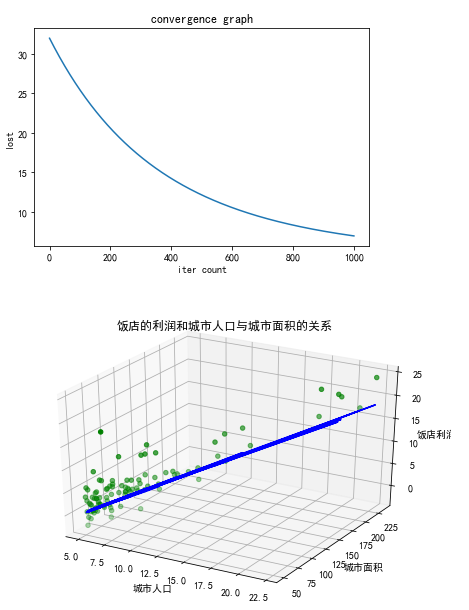


图 4

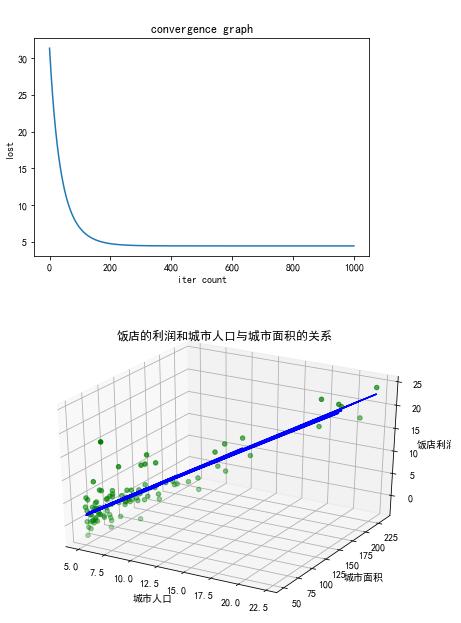


图 5