一轮考核

# 1.1 线性回归方程

### 1.1.1 思路

一开始，在考虑用梯度下降法还是最小二乘法，这两个学的时候有点模糊，用梯度下降法要迭代而且还要画图找学习率，后来看吴恩达的视频发现了一种叫正规方程的解法，看起来比较简单入手，（当时不知道他就是用最小二乘法解的）甚至还问了师兄才确定。然后当时（3.28）也只剩一周的时间，决定用最小二乘法做一个线性回归模型。

### 1.1.2 最小二乘法

#### 1.1.2.1推导

1)这是一个多特征值的线性回归方程，所以假设

**注：总体的矩阵斜率为： **

2）他的代价函数

3）m为样本总数，参数与特征矩阵X均为n+1维列向量

4）将代价函数写成矩阵表达式：

**注：X为m行n+1列的矩阵，为n+1维行向量，Y为m维行向量**

5）对以上公式进行变换，和求偏导，最终求出

6）所以，如果X是一个数据集时，将转置后最后的模型为

**注：此时的X是m（n+1）的矩阵，是（n+1）1的矩阵，得到的Y是一个m1的矩阵**

### 1.1.3 数据处理

首先，在jupyter notebook 把所有含有nan值的行都删去，最后只留下342行；

然后求出样本的每一个特征与outcome的相关系数，发现只有一组是低于0.5的；

接着，用scatter画出各个特征的离散图，看看每组数据的离散程度是怎样的，发现大部分是负相关，但有两组很怪，但是我只删掉了一组，就是相关系数过于低的“3”列所有特征值。

最后开始拿训练集来训练模型，得出后，拟合，然后再测试，得到最终结果。

参考文献：

[1](https://blog.csdn.net/qq\_37738656/article/details/83444764?msclkid=fc107758af7211ec9118900312aa9cf2) 正规方程的推导

[2] (<https://blog.csdn.net/A981012/article/details/105105318>) 不用框架去推导

[3] [【笔记】多元线性回归和正规方程解 - DbWong\_0918 - 博客园 (cnblogs.com)](https://www.cnblogs.com/jokingremarks/p/14283216.html) 最有用的一篇，教会我怎么封装类