

# 中山大学计算机学院

## 人工智能本科生实验报告

课程名称: Artificial Intelligence

教学班级	网安软工合班	专业 (方向)	网络空间安全
学号	20337251	姓名	伍建霖

### 一、实验题目

在给定文本数据集完成房价预测回归训练，画出训练loss曲线图、数据可视化图、预测函数图。

要求

- 设计合适的网络结构，选择合适的损失函数，利用训练集完成网络训练，画出数据可视化图、loss曲线图、预测函数图。
- 需要提交简要报告+代码
- 压缩包：学号\_姓名\_作业编号.zip，如 20331234\_张三\_实验7.zip。
- 截止日期：2022. 5. 25 23:59

### 二、实验内容

#### 1. 算法原理

单变量的代码助教给出了，所以下面讨论的都是多变量的情况

1. 先读入数据

2. 数据归一化

3. 用梯度下降法优化参数

#### 梯度下降法

1. 假定  $y = XW$

2. 设置损失函数并随机初始化网络参数  $W$

3. 当满足终止条件时，终止优化，否则继续

4. 计算  $y$  以及损失值

5. 求导可得梯度方向

6. 更新参数  $W$

7. 跳转到3

## 2. 伪代码

```
def regression:
    读入数据
    数据归一化
    g,cost=梯度下降法(x,y,theta,学习率,迭代次数)

def 梯度下降法:
    #初始化临时矩阵temp
    temp = np.matrix(np.zeros(theta.shape))
    #初始化代价函数举证，维数为迭代次数
    cost = []
    m = X.shape[0]
    for i in range(iters):
        #根据theta个数循环跟新theta
        temp = theta - (alpha/m)*(X*theta.T-y).T*X
        #得到跟新后的theta
        theta = temp
        #计算代价
        cost.append(computeCost(X, y, theta))
    return theta,cost
```

## 3.关键代码展示（带注释）

```
#求代价函数
def computeCost(X,y,theta):
    temp = np.power((X*theta.T-y),2)
    return np.sum(temp)/(2*X.shape[0])

#梯度下降
def gradientDescent(X,y,theta,alpha,iters):
    #初始化临时矩阵temp
    temp = np.matrix(np.zeros(theta.shape))
    #初始化代价函数举证，维数为迭代次数
    cost = []
    m = X.shape[0]
    for i in range(iters):
        #根据theta个数循环跟新theta
        temp = theta - (alpha/m) * (X*theta.T-y).T * X
        #得到跟新后的theta
        theta = temp
        #计算代价
        cost.append(computeCost(X, y, theta))
    return theta,cost

#标准化特征缩放
def preProcess(x,y):
    # 均值归一化
    x=(x-np.average(x))/(np.max(x)-np.min(x))
    x=np.c_[np.ones(len(x)),x]
    y=np.c_[y]
    return x,y
```

## 4.尝试

### 尝试更高的学习率

在单变量的情况下，0.01的学习率就足够生成一个可靠的预测模型，但在多变量的情况下，我使用0.01的学习率出现了cost越来越大的情况

### 不同的迭代次数

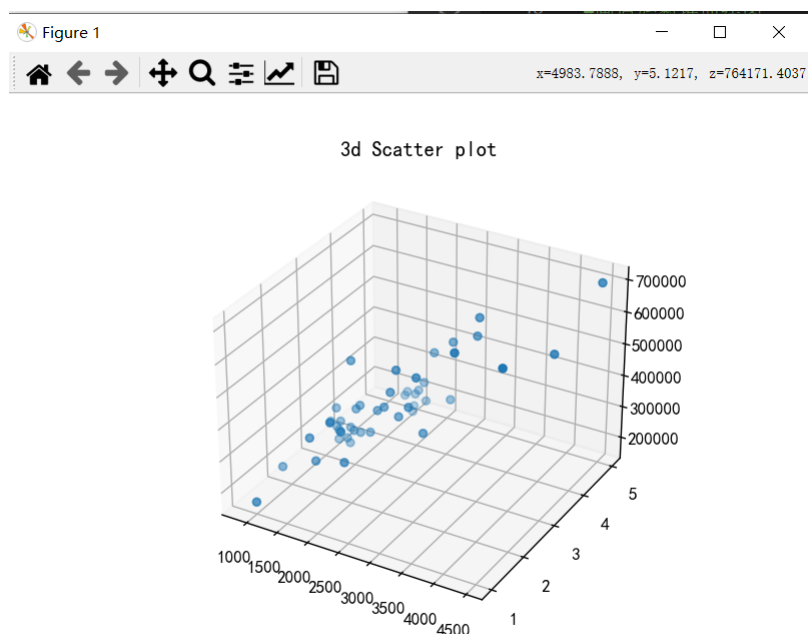
最终的迭代次数为1500，但从损失图像上看，cost在头几次的下降中就已经非常接近了最终的cost值(曲线就像个L型)，但我将iter缩小，参数又不够精准，出来的预测图与训练数据相差巨大

### 使用sklearn中现有的模型

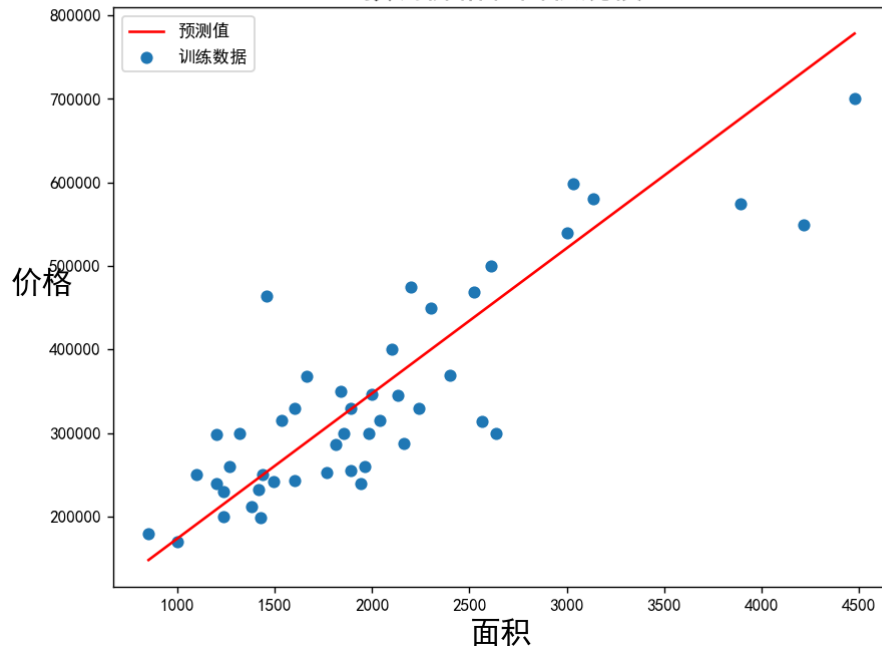
我手写的感知机模型最终的cost有7661236730(均方误差)，而使用sklearn中的linear\_model的LinearRegression模型fit后，最终cost仅有2140239948

## 三、实验结果及分析

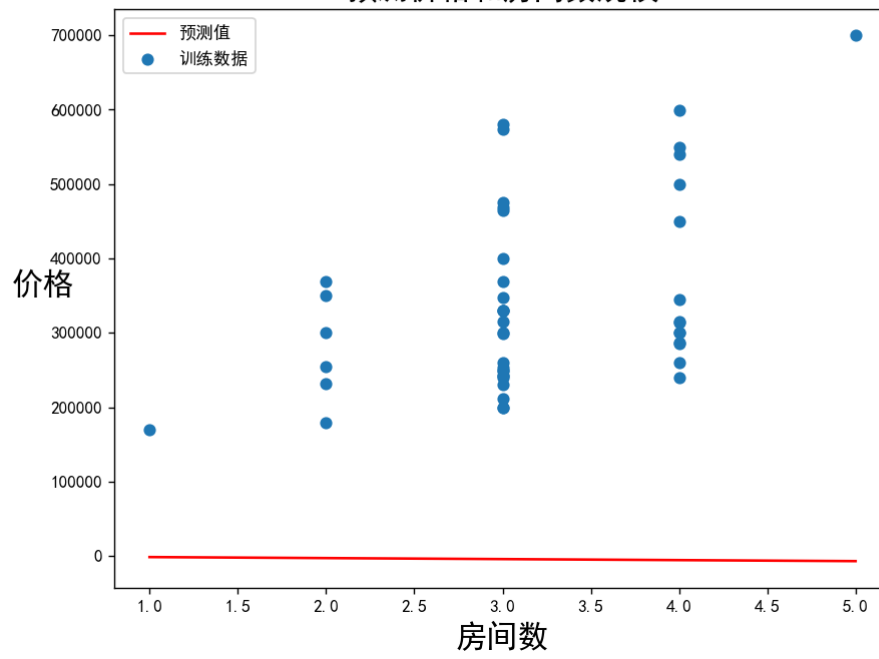
### 1. 实验结果展示示例（可图可表可文字，尽量可视化）

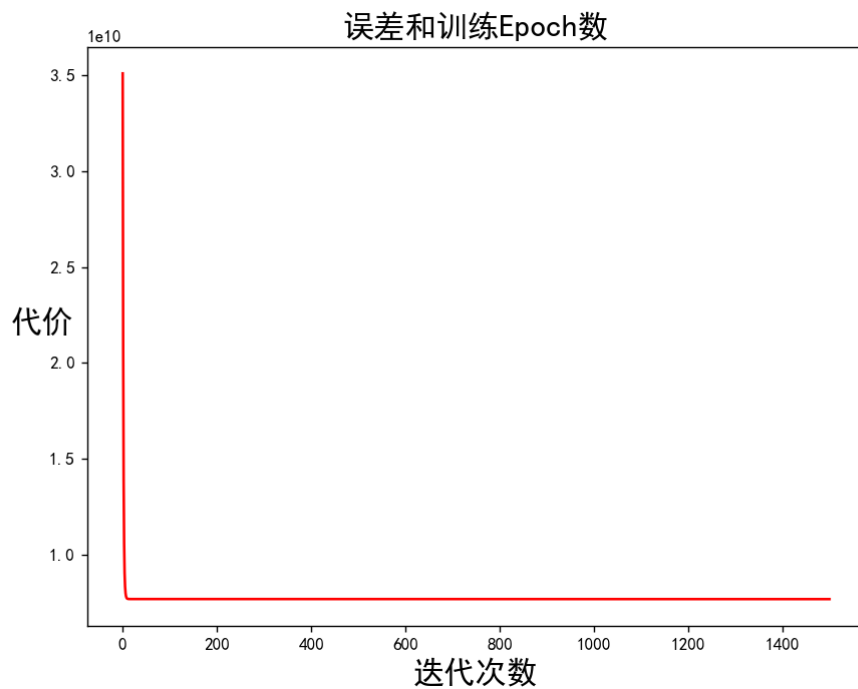


预测价格和面积规模



预测价格和房间数规模





## 2.评测指标展示及分析（其它可分析运行时间等）

面积和价格函数的斜率：173.78260468148616。

房间数和价格函数的斜率：-1362.9672962825705。

最终cost：7661236730.926858。

## 3.对模型的一些思考

生成预测图后，我对房间数和价格函数的斜率，以及最终cost有些不解。我试着给出解释：可能房间数对价格的影响微乎其微，但至于为什么斜率为负数，这我百思不得其解；至于最终上十位数的cost，我认为应该是均方误差的原因，毕竟其中有个平方存在

## 四、参考资料

理论课和实验课PPT