

## 实验三 线性回归

### 一、一元线性回归

#### 1、实验描述

1) 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。

2) 学习一元线性回归的损失计算。

3) 使用批梯度下降法。

4) 使用matplotlib画出回归曲线

5) 定义predict实现预测功能

#### 2、实验及结果分析

(1) 开发语言及运行环境;

python3.9.19, Mindspore2.4.1。

(2) 实验的具体步骤;

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)

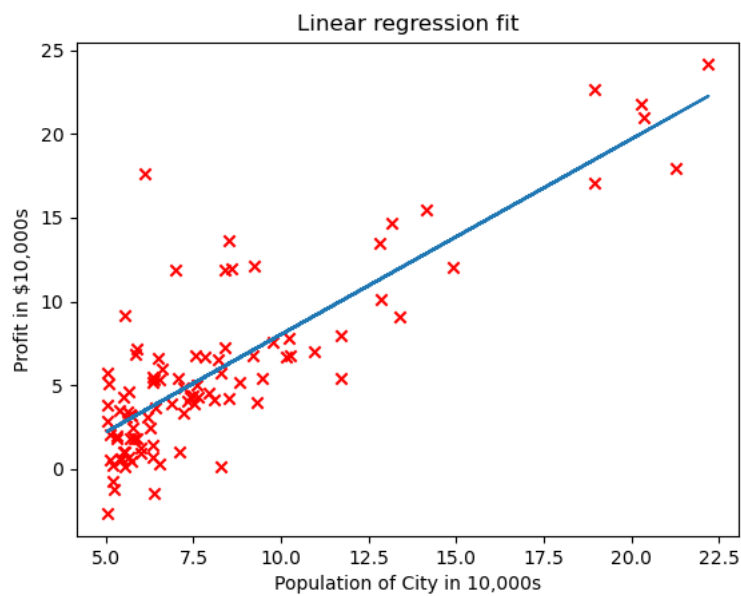
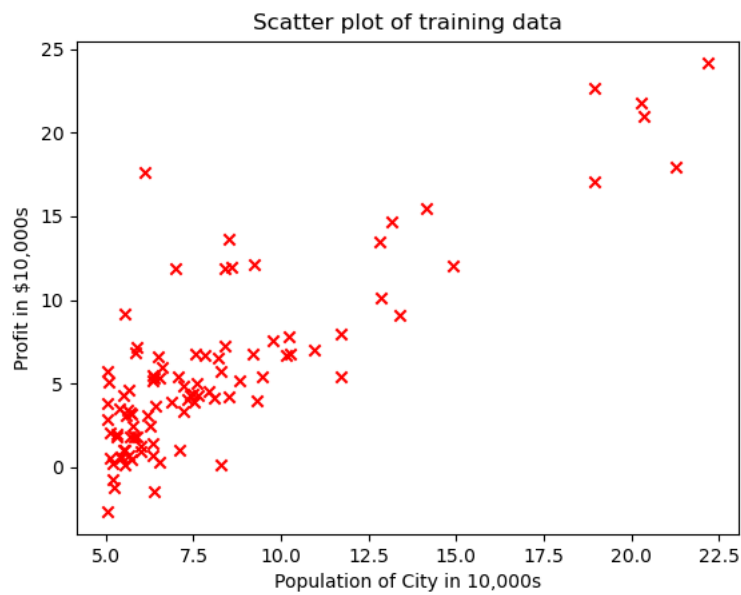
2) 将线性回归参数初始化为0, 然后计算代价函数 (cost function) 并求出初始值。

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。

(3) 根据实验数据集，按实验要求给出相应的结果（截图）；



初始代价：32.072733877455676

参数向量：[-3.63029144 1.16636235]

人口为35000时，收益为 4519.7678677017775

人口为70000时，收益为 45342.45012944715

■

(4) 对实验结果进行简要分析。

根据拟合图形，一元线性回归的拟合效果并不是很理想。

### 3、实验心得

(1) 首先通过散点图展示了城市人口与移动餐车利润之间的关系。图中显示了数据点的分布情况，直观地了解数据的特征。

(2) 将线性回归参数初始化为0，并计算了初始代价函数值。初始代价函数值较高，表明初始参数并不能很好地拟合数据。

(3) 使用梯度下降法对线性回归参数进行了优化。经过1500次迭代，参数向量 $\theta$ 收敛到了一个较优值。

### 5、程序文件名的清单

源码/exp3\_1.py

## 二、多元线性回归

### 1、实验描述

题目：

1) 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。

2) 数据加载和归一化，将特征和目标值转换为MindSpore的Tensor格式

3) 使用均方误差（MSE）作为损失函数。

4) 使用Adam优化器进行梯度下降，更新模型参数。尝试不同的学习率，选择最优的学习率。

5) 使用训练好的模型和归一化的特征进行预测。

数据集：

文件ex1data2.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。

## 2、实验及结果分析

(1) 开发语言及运行环境；

python3.9.19, Mindspore2.4.1。

(2) 实验的具体步骤；

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，对进行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。

(3) 根据实验数据集，按实验要求给出相应的结果（截图）；

**最佳学习率：0.01，最小损失：8.87479e-05**

**预测房价：293081.50**

(4) 对实验结果进行简要分析。

由于测试数据较少，训练循环取400次时，最佳学习率为0.01。

## 3、实验心得

(1) 学习使用计算标准差和均值，对特征进行归一化。

(2) 学会了计算线性回归方程的损失函数计算。

(3) 学会了使用Mindspore训练简单的人工神经网络

## 5、程序文件名的清单

源码/exp3\_2.py