

西安电子科技大学

机器学习实验报告



姓名	鲍超俊
学号	15020510059
指导老师	杨淑媛

- 一、实验内容
- 二、实验原理（线性回归）
 - 1 算法简介
 - 2. 线性回归模型
 - 2.1 问题描述
 - 2.2 模型建立
 - 2.3 优化目标
 - 3. 优化过程
- 三、实验结果
- 四、程序说明
 - 1. 工具包清单
 - 2. 模块功能

一、实验内容

The task is to write a program in MATLAB to fit a straight line in linear regression .Use the steepest gradient descent to perform loss function. The data being used have been given. Please give out the loss function decline curve, the iteration number, and the final fitting line .

二、实验原理（线性回归）

1 算法简介

在统计学中，线性回归(Linear Regression)是利用称为线性回归方程的最小平方函数对一个或多个自变量和因变量之间关系进行建模的一种回归分析。这种函数是一个或多个称为回归系数的模型参数的线性组合。只有一个自变量的情况称为简单回归，大于一个自变量情况的叫做多元回归。

2. 线性回归模型

2.1 问题描述

若已知 m 个样本，每个样本对应于 n 维特征和一个结果输出，如下：

$$(x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)}, y_0), (x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}, y_1), \dots, (x_1^{(m)}, x_2^{(m)}, \dots, x_n^{(m)}, y_n)$$

那么，对于一个新的 $(x_1^{(x)}, x_2^{(x)}, \dots, x_n^{(x)})$ ，预测其对应的 y_x 。

2.2 模型建立

假设 x 与 y 满足线性关系，不妨设预测值为 h_θ ：

$$h_\theta(x_1, x_2, \dots, x_n) = \theta_0 + \theta_1 x_1 + \dots + \theta_n x_n$$

其中 θ_i 为模型参数。上式也可简化成矩阵形式：

$$h_\theta(x) = x \cdot \theta$$

2.3 优化目标

不妨定义损失函数如下：

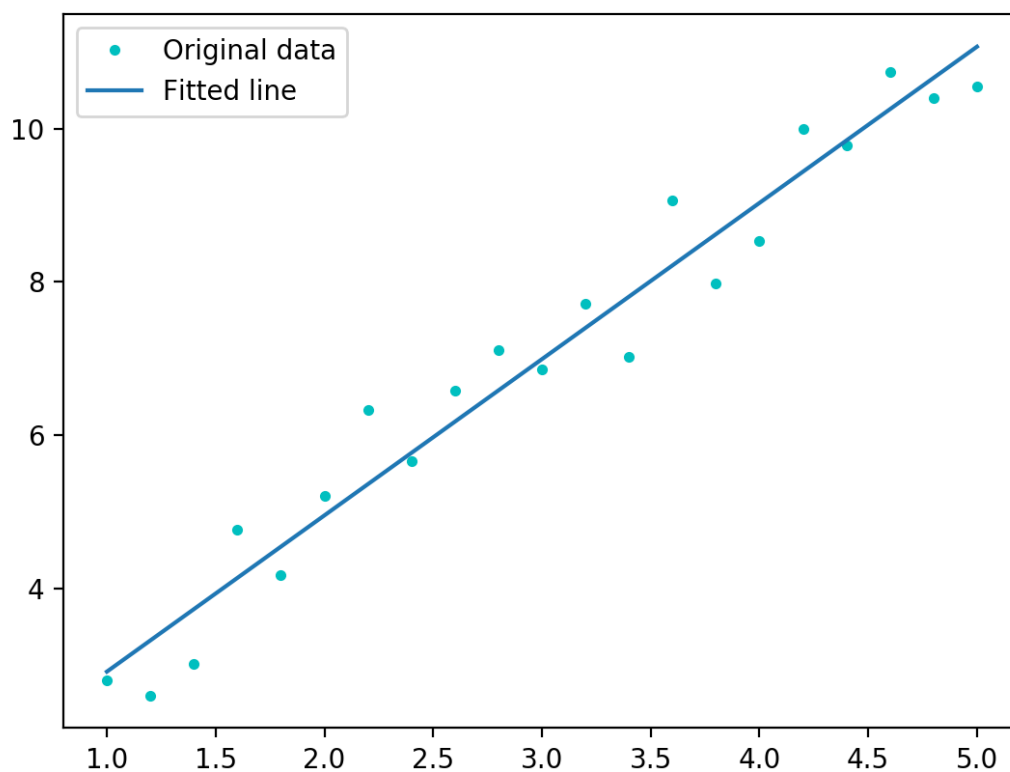
$$J(\theta_0, \theta_1, \dots, \theta_n) = \sum_{i=0}^m (h_\theta(x_0, x_1, \dots, x_n) - y_i)^2$$

3. 优化过程

对于线性回归的损失函数常用梯度下降法来优化。 θ 的更新公式如下：

$$\theta = \theta - \alpha X^T (X \cdot \theta - Y)$$

三、实验结果



四、程序说明

1. 工具包清单

- tensorflow
- numpy
- matplotlib.pyplot

2. 模块功能

main.py

```
def import_data():  
    """  
    从./data.csv 导入数据  
    :return: [x,y]  
    """  
    pass
```

```
def train(X, max_iter=4000, learn_rate=1e-2):  
    """  
    训练模型  
    :param X: 数据  
    :param max_iter: 最大迭代次数, 默认6000  
    :param learn_rate: 梯度下降步长, 默认1e-2  
    :return:  
    """  
    pass
```

```
def draw(weight, bias, X):  
    """  
    画图  
    :param weight: 权重  
    :param bias: 偏置  
    :param X: 样本  
    :return:  
    """  
    pass
```