**机器学习**

**作业：线性回归**

**最后期限： 周一 10/17 晚上 11:59 (CDT)**

## 一、宗旨：

本作业包含线性回归和逻辑回归的经典回归任务，是本课的预热项目。 请使用 Google **Colab完成此作业**。

* 了解如何使用CoLab 。
* 了解如何加载自定义数据。
* 了解如何对线性回归执行梯度下降。
* 了解如何绘制中间结果。

二。下载数据

food\_truck\_data.txt： [https://drive.google.com/file/d/1J\_xP-bW\_vVQIFRnE0ApH76jbaY51Zqkv/view ?usp=sharing](https://drive.google.com/file/d/1J_xP-bW_vVQIFRnE0ApH76jbaY51Zqkv/view?usp=sharing)

housing\_price\_data.txt:

<https://drive.google.com/file/d/1LWOJrECQbYyptGwhKBLXL_z0hGONX9Rb/view?usp=sharing>

## 二。描述

任务 1：单变量线性回归

假设您是一家连锁餐厅的首席执行官，正在考虑在不同城市开设一家新店。该连锁店已经在各个城市设有快餐车，并且您拥有城市的利润和人口数据。您想计算出一辆新快餐车的预期利润是多少，只考虑它所在城市的人口。根据可用数据使用简单回归预测一条快餐车的利润。

文件food\_truck\_data.txt包含问题的数据集。数据由两列组成；第一列是一个城市的人口，第二列是该城市食品卡车的利润。利润的负值表示亏损。

* 将线性模型拟合到数据（有关详细的评分术语，请参阅下面的第 III 部分。评分和提交）
  + 使用梯度下降计算模型的参数
  + 在执行梯度下降迭代时计算并绘制成本函数
* 使用线性模型绘制数据（有关详细的评分术语，请参阅下面的第 III 部分。评分和提交）

任务 2：多变量线性回归

假设你正在卖房子，你想知道一个好的市场价格是多少。您可以使用有关近期房屋销售价格的可用数据并学习线性模型。

文件housing\_price\_data.txt包含问题的数据集。数据由三列组成；第一列是房子的大小（以平方英尺为单位），第二列是卧室的数量，第三列是房子的价格。

* 将线性模型拟合到数据（有关详细的评分术语，请参阅下面的第 III 部分。评分和提交）
* 具有以下要求（详细评分条款请见下文第三部分评分和提交）
  + 使用特征缩放
  + 使用梯度下降计算参数
  + 在执行梯度下降迭代时计算并绘制成本函数

## 三、评分和提交

* 该作业将以总计7 0 分的形式进行评估。一般根据下表给出基本分数。

|  |
| --- |
| 使用 Google Colab （10 分）   * 提交工作Colab ipynb文件 (5') * 将所需结果嵌入 ipynb ( 5')   第 1 部分：单变量线性回归（ 3 0 分）   * 为数据定义线性模型 (5') * 在文本单元格中，用 >=50 个词解释您定义的模型（例如，哪些参数是输入，哪些是输出，哪些是要计算的） ( 5' ) * 使用显式梯度下降计算模型的参数 ( 5' )   \*请不要使用封闭式解决方案或非交互式解决方案。   * 梯度下降形成为迭代循环 (5') * 在执行梯度下降迭代时绘制成本函数 (5') * 使用线性模型绘制数据 (5')   第二部分：多变量线性回归（ 3分0分）   * 使用特征缩放 (5') * 为数据定义线性模型 (5') * 在文本单元格中，用 >=50 个词解释您定义的模型（例如，哪些参数是输入，哪些是输出，哪些是要计算的）(5') * 使用显式梯度下降计算模型的参数 (5')   \* 请不要使用封闭式解决方案或非交互式解决方案。   * 梯度下降形成为迭代循环 (5') * 在执行梯度下降迭代时绘制成本函数 (5') |

对于每个 5' 刻度。

5' = 完全正确

4' = 小瑕疵

3' = 大部分不正确

2' = 完全不正确

1' = 做了某事

0'= 什么都没做

附录（一些注释）

1. 本作业的示例代码

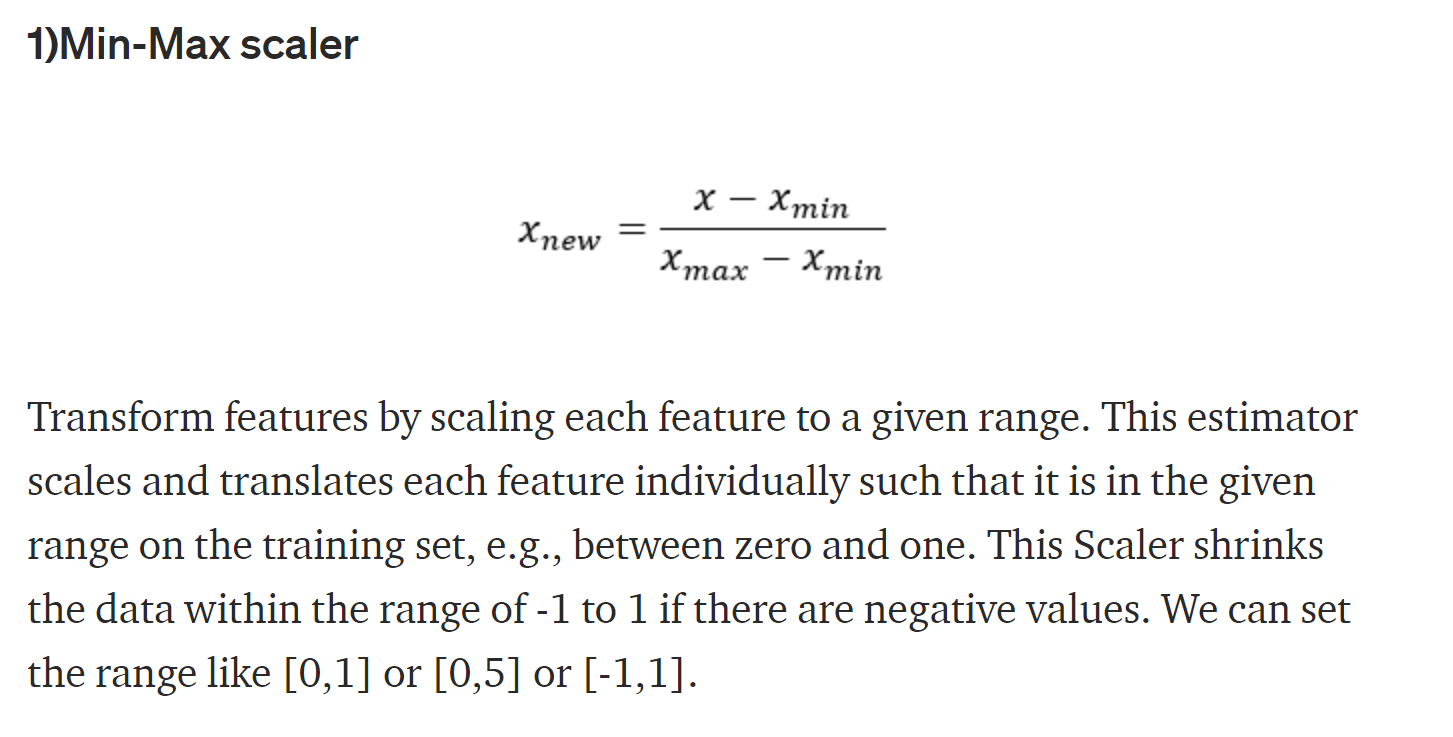
<https://www.cs.toronto.edu/~frossard/post/linear_regression/>

你可以参考这个实现。 （但请不要复制粘贴代码）

1. 对于特征缩放，多变量数据集中的输入范围变化很大。所以我们使用特征缩放来规范化输入。特征缩放有很多不错的选择

<https://towardsdatascience.com/all-about-feature-scaling-bcc0ad75cb35>

最简单的是最小-最大缩放器



1. 成本曲线是成本迭代图，比如

