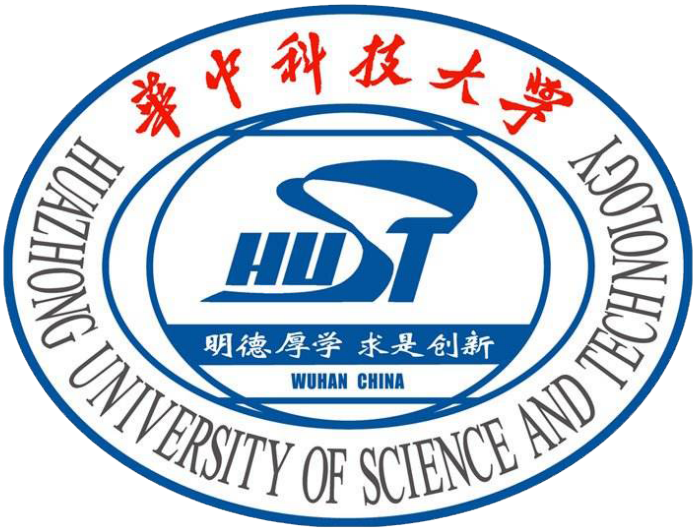
华中科技大学计算机科学与技术学院

《机器学习》课堂三结课报告



专　　业	计算机科学与技术
班　　级	ACM1901
学　　号	U201915010
姓　　名	李恒庄
成　　绩	
指导教师	何琨
时　　间	2021 年 11 月 28 日

目录

1 实验题目 基于差分进化算法的多段线性回归

1.1 摘要

1.2 多段线行回归问题介绍

选择这个题目的源起是上了何琨老师的一节算法课后，何老师给我们留了一道作业题：考虑一个如图 1.1 所示的数据集，我们可以观察到可以用多段线性函数来拟合这个数据集，我们需要确定将该数据集分为多少个聚类，并找到相应断点 (*breakpoint*) 使得代价函数 J（这里的代价函数可以有多种定义的方法）最小:

$$J(\boldsymbol{a}, \boldsymbol{b}, k) = \frac{1}{N} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (a_i \cdot x_j^{(i)} + b_i - y_j^{(i)})^2$$

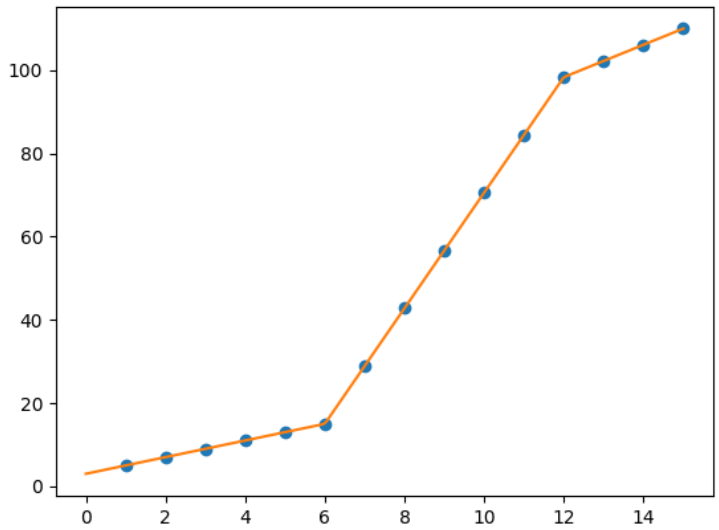


图 1.1: 多段线性回归

这道题乍看上去似乎很简单，不就是分段线性回归吗？但是当我课下实际下手做，发现事情并没有我想象的那么简单。粗略地查找相关文献之后，我发现断点数以及断点位置的确定仍然是解决这一问题的难点，很多学者都给出了解决办法，但是很多都是基于一定的数据特征以及实际应用限制，很少有给出一种普适而简约算法的。

由此，我发现研究此道题目对本科水平同学是一种有益的探索。所以我决定以这道题目作为我的机器学习课程的研究对象。

事实上，多段线性回归问题在近十几年都是一个比较活跃的问题，原因在于其所得拟合函数非常简单，对于一些轻量化或者定性问题的研究中都是时分有益的；在实际工

业生产中，譬如滤波器设计等，面对繁杂的代价函数，人们更希望得到的函数是简单的；面对潜在信息未知的问题，多段线行回归能够在快速地给出对研究对象的粗略解释，方便人们掌握其中规律。

1.3 差分进化算法

在不断地检索文献后，我发现一个经典算法——差分进化算法 (Differential Evolution Algorithm, DE) 能够用于解决这道问题中断点应该设置在哪里的问题（当得到断点后，问题给就变得非常简单了）。所以我决定以差分进化算法作为主体探索解决多段线行回归问题的方法。差分进化算法在上世纪由 Storn 等人提出，接下来便被证明是收敛最快的进化算法，在接后的二十多年里不断被改进。

1.4 研究现状

多段线性回归早在本世纪初就已经在实际应用中被研究，从最开始的一维变量线性回归，经过无数学者的发展和研究，逐步扩展到多维回归问题以及多段非线性回归问题。相关研究从局限于小规模问题到大规模问题，从串行计算到并行分布式计算。同时，多段线性回归还被用来构建专家系统，是这一传统领域又有了新的突破；多段线行回归方法的引入可以使专家系统更加有效的获取所需信息，譬如可以使得医学中肿瘤评估的工作效率得到极大提高，通过多段线行回归改进的专家系统给出复杂医学归纳问题的答案。

限于篇幅，研究现状不再赘述。同时，本实验报告的多个部分均限于篇幅无法给出详细的证明过程、思维过程等，仅记载我大致的摸索过程以及遇到的相关问题和解决办法。

2 实验要求

1. 自行从多种数据分布中采样构建不同数据集，包括线性决策边界和各种非线性决策边界；
2. 自行划分训练集与测试集；
3. 使用课程中学习过的分类器进行分类预测，包括核方法实现；
4. 评估不同模型在不同数据集上的表现，并给出相应分析及思考；
5. 自主拓展探索；
6. 严禁直接调用已经封装好的各类机器学习库（包括但不限于 sklearn），但可以用 NumPy 等数学运算库，严禁抄袭网上代码。

3 算法设计

3.1 环境

3.2 系统功能需求

3.3 系统设计

3.4 系统实现

3.5 系统测试及结果说明

3.6 其他需要说明的问题

4 实验环境与平台

4.1 环境

4.2 系统功能需求

4.3 系统设计

4.4 系统实现

4.5 系统测试及结果说明

4.6 其他需要说明的问题

5 程序实现

5.1 环境

5.2 系统功能需求

5.3 系统设计

5.4 系统实现

5.5 系统测试及结果说明

5.6 其他需要说明的问题

6 实验结果

6.1 环境

6.2 系统功能需求

6.3 系统设计

6.4 系统实现

6.5 系统测试及结果说明

6.6 其他需要说明的问题

7 结果分析

7.1 环境

7.2 系统功能需求

7.3 系统设计

7.4 系统实现

7.5 系统测试及结果说明

7.6 其他需要说明的问题

-

8 机器学习课程的学习体会与建议

8.1 环境

8.2 系统功能需求

8.3 系统设计

8.4 系统实现

8.5 系统测试及结果说明

8.6 其他需要说明的问题

参考文献

[1] James F. Kurose, Keith W. Ross. 计算机网络: 自顶向下方法 (第 7 版) [M]. 机械工业出版社, 2018.

A 附录 I