

计算机与信息技术学院

本科生《机器学习》课程大作业

实验报告

**实验名称 分类与回归任务有监督模型对比分析**

**小组成员 17292012 何翔**

**17292008 王家鹏**

**日 期 2020.11.27**

**一、实验内容**

* 1. 分类任务有监督模型综合对比分析

分类任务有监督模型综合对比分析需要同学们自行搜集分类任务的数据集（一个即可），使用学过的逻辑回归、决策树、神经网络、支持向量机四个模型（可以使用Scikit-learn或其他工具包）在该数据集上进行对比评测，要求：

1. 每个方法尝试不同参数，运用Accuracy, Precision, Recall, F1四个指标进行对比评测，把结果绘制成表格（四个模型四个表格）和可视化图（只可视化数值型参数即可，如SVM的C，决策树的深度等，每个模型每个参数一张图）
2. 运用第1条尝试得到的最优参数对四个模型进行横向对比，计算Accuracy, Precision, Recall, F1四个指标，把结果绘制成表格

* 2. 回归任务有监督模型综合对比分析

回归任务有监督模型综合对比分析需要同学们自行搜集回归任务的数据集（一个即可），使用学过的线性回归、决策树、神经网络、支持向量机四个模型（可以使用Scikit-learn或其他工具包）在该数据集上进行对比评测，要求：

1. 每个方法尝试不同参数，运用MSE,RMSE,MAE三个指标进行对比评测，把结果绘制成表格（四个模型四个表格）和可视化图（只可视化数值型参数即可，如SVM的C，决策树的深度等，每个模型每个参数一张图）
2. 运用第1条尝试得到的最优参数对四个模型进行横向对比，计算MSE,RMSE,MAE三个指标，把结果绘制成表格

**二、实验设计**

* 1. 分类任务有监督模型综合对比分析

选择逻辑回归、决策树、神经网络、支持向量机四个模型在heart-disease上的分类，运用Accuracy, Precision, Recall, F1四个指标进行对比评测

* 2. 回归任务有监督模型综合对比分析

选择逻辑回归、决策树、神经网络、支持向量机四个模型在Abalone上的回归，运用MSE,RMSE,MAE三个指标进行对比评测

**三、实验环境及实验数据集**

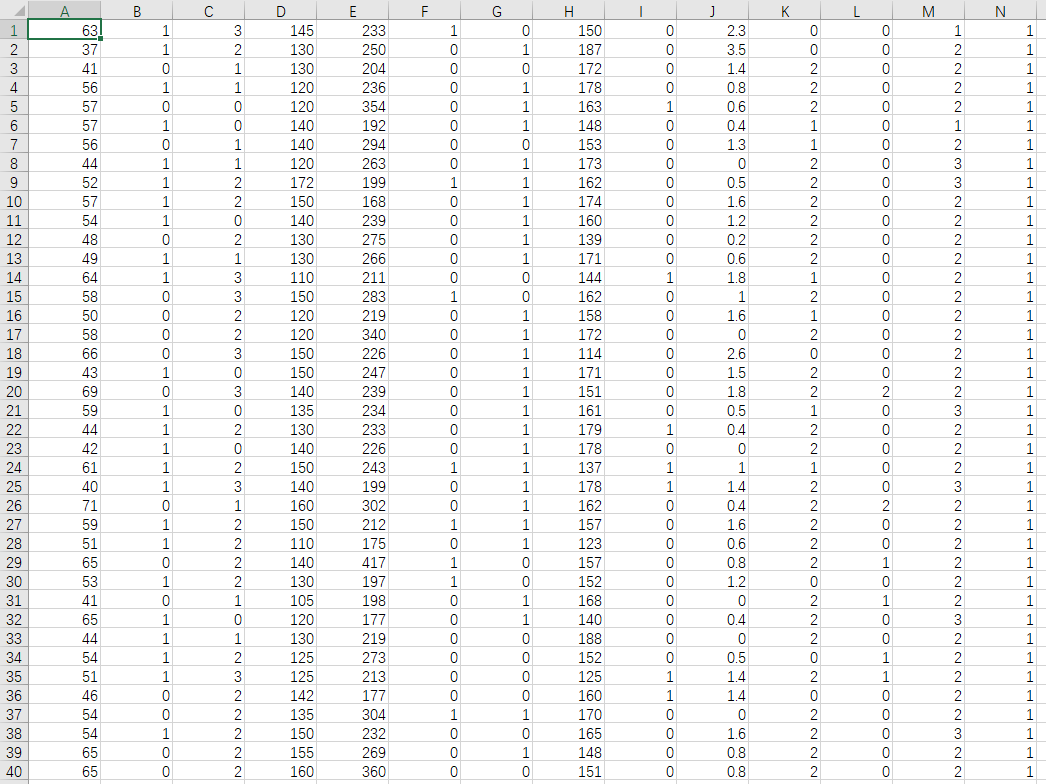
操作系统：Win10

Python版本：3.6.12

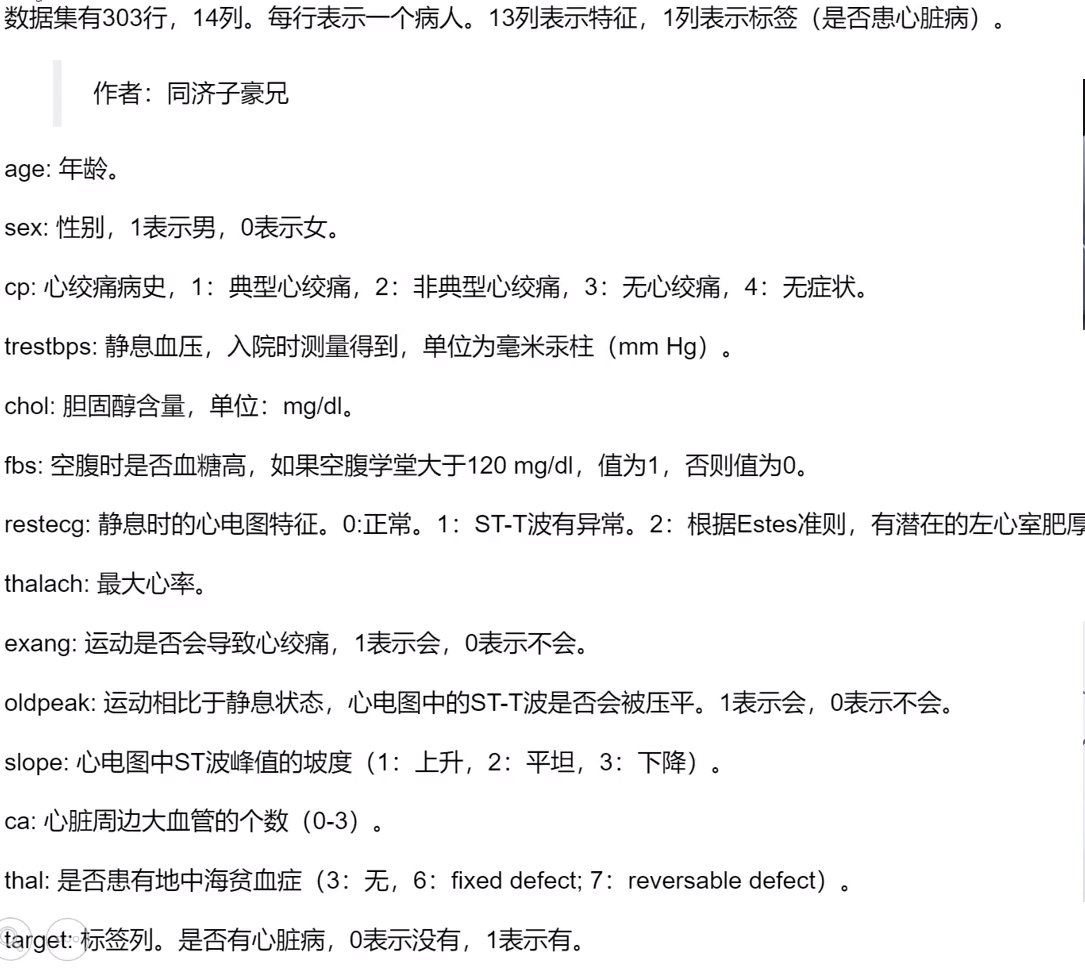
CPU：i5-4590 16G

数据集heart-disease：<https://archive.ics.uci.edu/ml/datasets/Heart+Disease>

数据集内容：



数据集说明：



数据集Abalone：<https://archive.ics.uci.edu/ml/datasets/Abalone>

此数据集数据以逗号分隔，没有列头。每个列的名字存在另外一个文件中。

给出了属性名称，属性类型，度量单位和简要说明。环数是要预测的值：可以是连续值，也可以是分类问题。

名称/数据类型/度量单位/描述

-----------------------------

性别/名义/-/ M，F和I（婴儿）

长度/连续/毫米/最长的外壳尺寸

直径/连续/毫米/垂直于长度

高度/连续/毫米/带有外壳的肉总

重/连续/克/整个鲍鱼

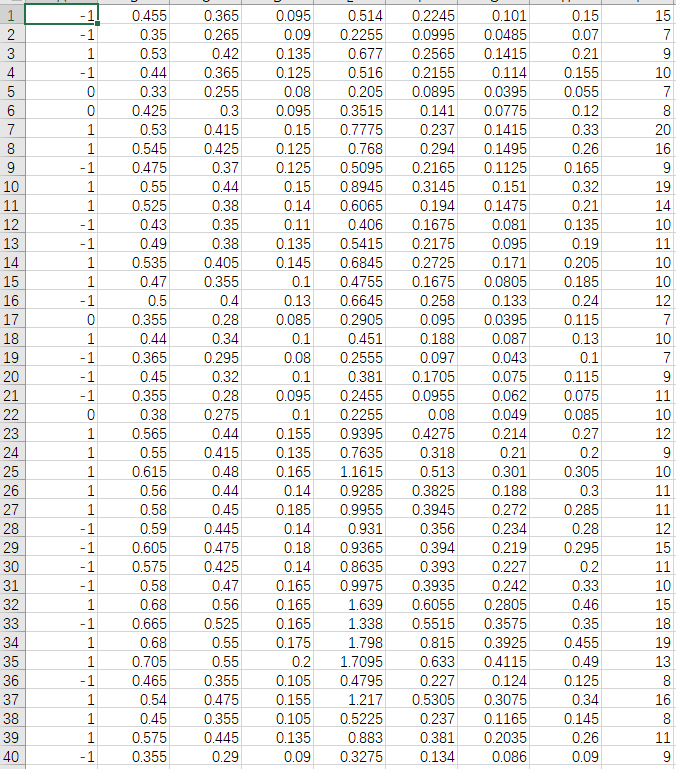
去皮重量/连续/克/重量肉

内脏重量/连续/克/肠重量（出血后）

蛋壳重量/连续/克/干后

环/整数/-/ +1.5给出年龄（以年为单位）

将性别male,female,infant做-1,1,0的处理，内容如下：

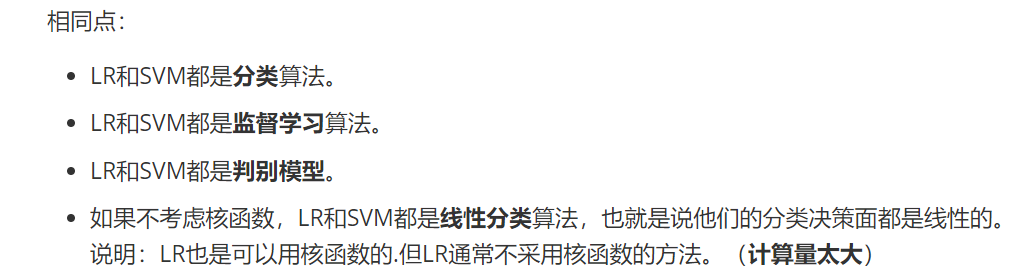


**四、实验原理**

* 1. 模型评估



* 2.逻辑回归与SVM的异同



* 3.常用分类算法的优缺点



**五、实验过程与结果**

* 1. 分类任务有监督模型综合对比分析

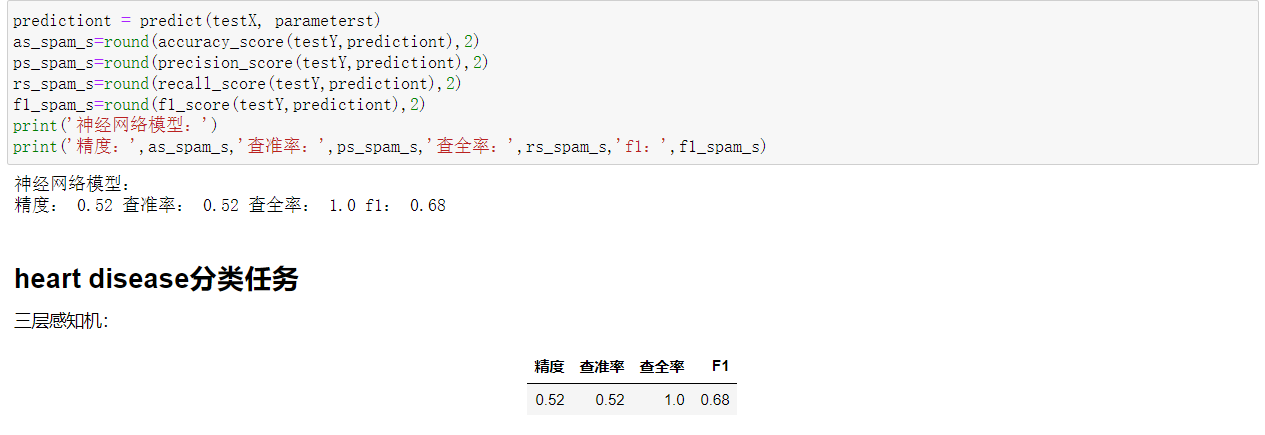
1. Logistics



1. 决策树



1. 神经网络



4）SVM



5）模型的横向对比

运用各个模型的最优参数对四个模型进行横向对比

Logistics最大迭代max\_iter=100

决策树最大深度为3

神经网络三层感知机

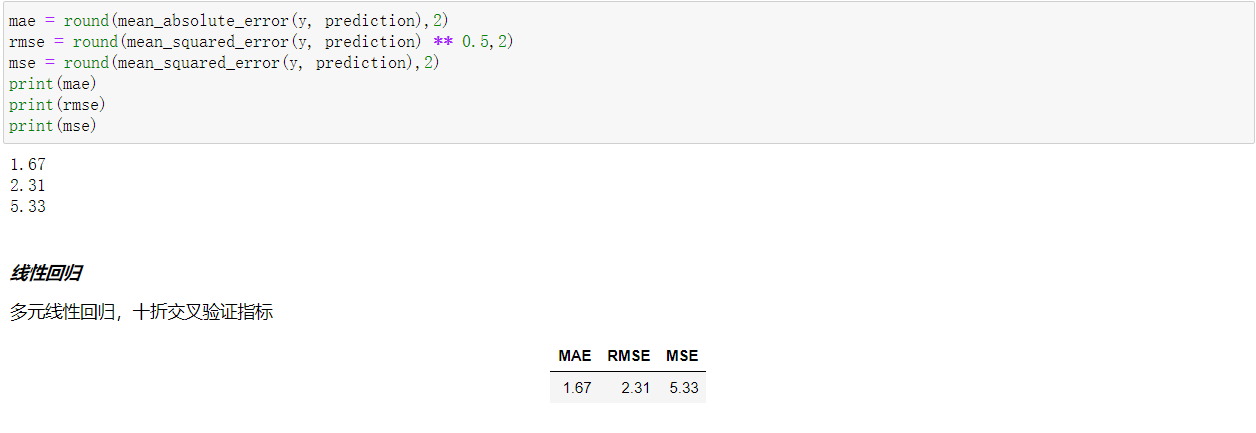
SVM的linear核函数C=0.7



* 2. 回归任务有监督模型综合对比分析

选择逻辑回归、决策树、神经网络、支持向量机四个模型在Abalone上的回归，运用MSE,RMSE,MAE三个指标进行对比评测

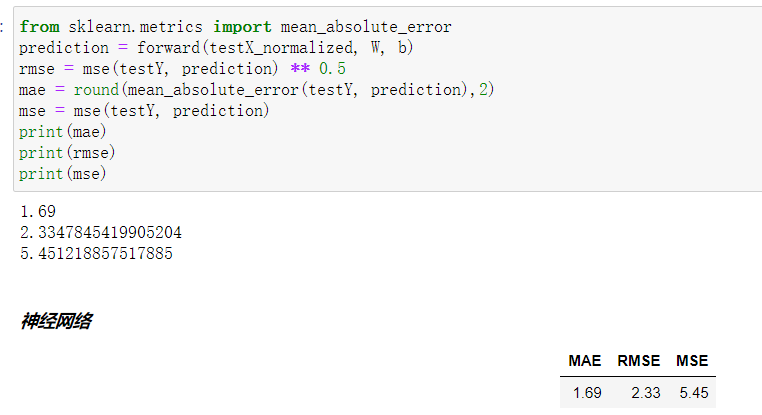
1. Logistics



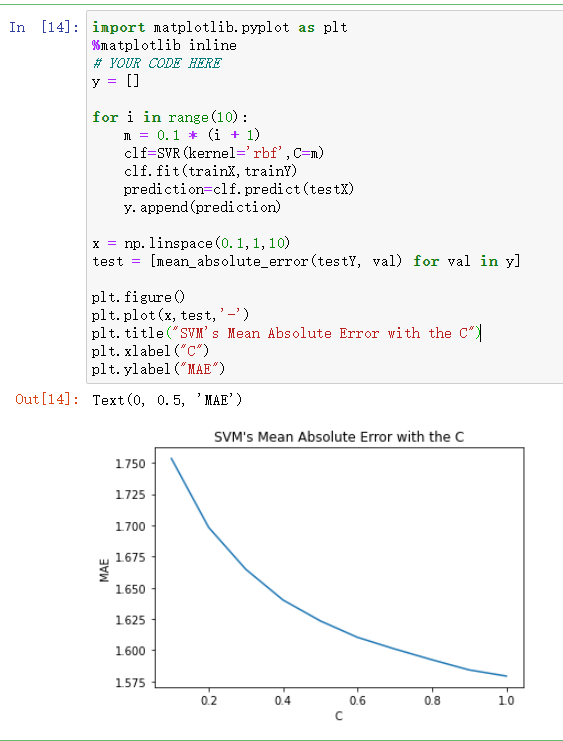
1. 决策树



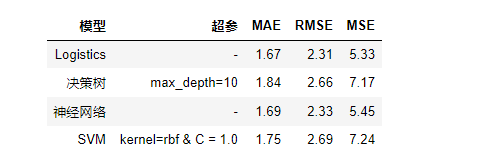
3）神经网络



4）SVM



5）模型的横向对比



* 3. 对比总结

对模型的横向对比可以发现，Logistic Regression简单易于理解，能容易地更新模型吸收新的数据。SVM的小样本机器学习受欢迎，泛化性能比较高。但其运行时间和超参的调整略烦人。决策树能在短时间内处理大量数据，得到可行且效果较好的结果，但容易忽略属性之间的相关性，造成过拟合。神经网络的并行处理能力强，但结果难以解释。

**六、实验心得体会**

监督学习可概括如下：从给定有限的数据出发，假设数据是独立同分布的，而且假设模型属于某个假设空间，应用某一评价准则，从假设空间中选取一个最优的模型，使它对已给训练数据及未知测试数据在给定的评价标准意义下有最准确的预测。本次实验中从逻辑回归、决策树、神经网络、支持向量机四个模型系统地进行了分类与回归任务有监督模型对比分析。通过实验资料的学习和程序的编写运行，有监督的学习有了直观的认识。自己动手查找数据集并根据特定需求对其进行修改也是必不可少的环节。

**七、参考文献**

【1】鲍鱼数据集处理/matlab

https://blog.csdn.net/xingdu\_/article/details/106896478

【2】DeepLearning-500-questions

https://github.com/XDUSPONGE/DeepLearning-500-questions

**八、附录**

见文件中8个源代码