

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**

**小组成员**

**日 期**

1. **实验环境**

#### ···· 学习框架：anaconda3（Win10），数据集：sklearn.datasets.load\_boston() 数据集。

1. **线性模型**
2. **多元线性模型**

在多元线性模型中，使用全部特征，在十折交叉验证下得到MAE、MSE和RMSE指标评测如下。



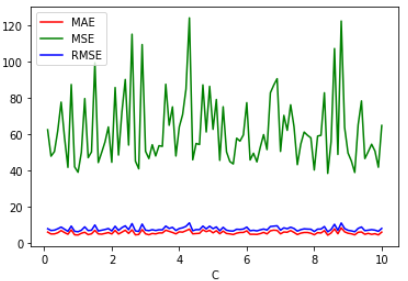
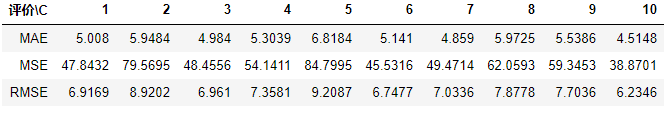
1. **多元逻辑回归模型**

在多元逻辑回归模型中，使用全部特征，在十折交叉验证下得到MAE、MSE和RMSE指标评测如下。



1. **SVM**
2. **LinearSVR**

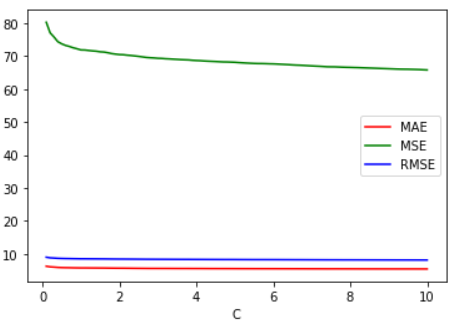
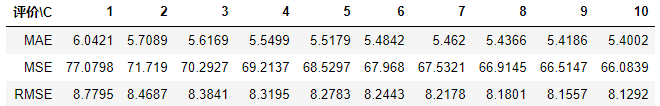
使用全部特征，线性SVM回归中C从0.1~10，步长为0.1，在十折交叉验证下得到MAE、MSE和RMSE指标评测如下图所示。在下表中截取1~10，间隔为1的指标评测。



线性SVM回归中，在MAE、MSE和RMSE指标下评测，C值对性能影响是波动的，在此数据集下取C=1，效果比较好。

1. **SVR-rbf**

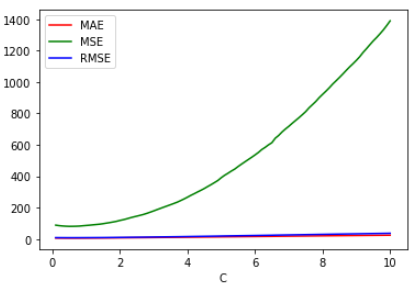
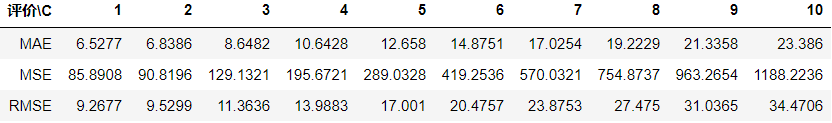
使用全部特征，使用rbf核。C的值从0.1~10，步长为0.1，在十折交叉验证下得到MAE、MSE和RMSE指标评测如下图所示。在下表中截取1~10，间隔为1的指标评测。



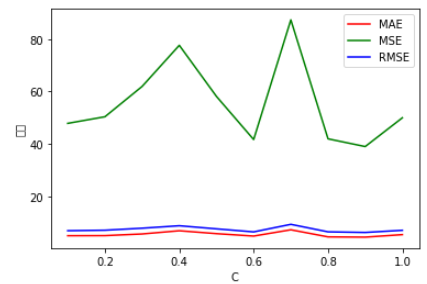
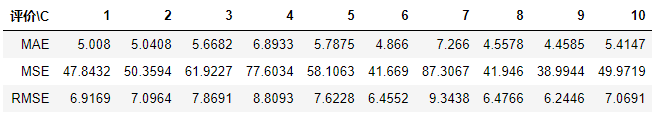
SVR-rbf的性能在MAE、MSE和RMSE指标下评测，随着C的增加其性能不断提高，此数据集下取C=8，效果较好。

1. **SVR-sigmoid**

使用全部特征，使用sigmoid核。C的值从0.1~10，步长为0.1，在十折交叉验证下得到MAE、MSE和RMSE指标评测如下图所示。在下表中截取1~10，间隔为1的指标评测。



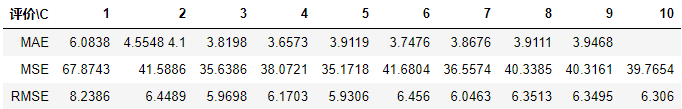
由上图可知，C在0.1-1之间有一个最优值，打印0.1~1的值如下：



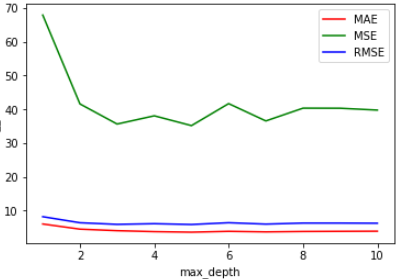
SVR-sigmoid的性能在MAE、MSE和RMSE指标评测下，随着C的增加先提高再下降：在此数据集下取C=0.6，效果较好.

1. **决策树**

调整决策树的深度（1~10），在十折交叉验证下得到MAE、MSE和RMSE指标各个值如下表所示：



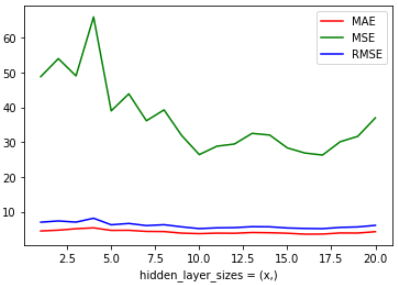
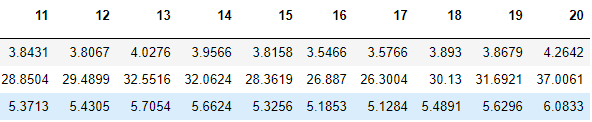
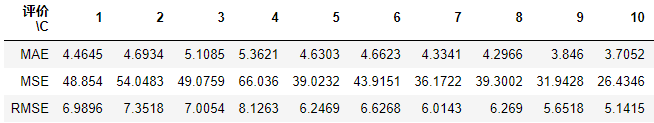
评测如下图所示：



在决策树模型中，MAE、MSE和RMSE指标评测下，最大深度设置为3时MAE、MSE和RMSE指标下评测较好

1. **神经网络**
2. **一个隐含层**

只有一个隐含层，而输入层只有13个神经元，隐含层神经元个数尝试1~20

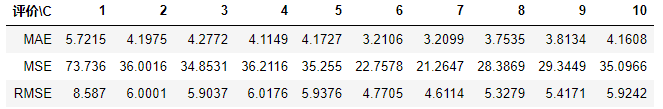
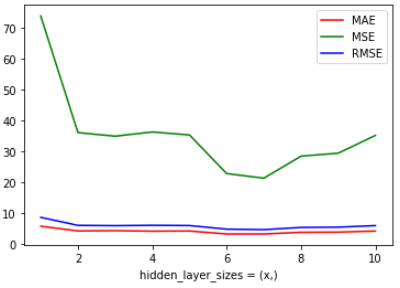


只有一层隐含层时，在MAE、MSE和RMSE指标下评测下，有10个神经元时效果最好。

1. **两个隐含层**

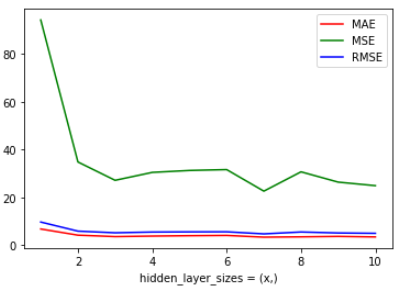
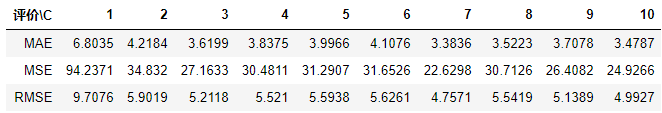
在实验中，第一个隐含层拥有个5~11个神经元，在第一层确定的情况下，第二层的神经元个数从1到10，使用十折交叉验证逐个实验.

1. 第一个隐含层拥有个5神经元



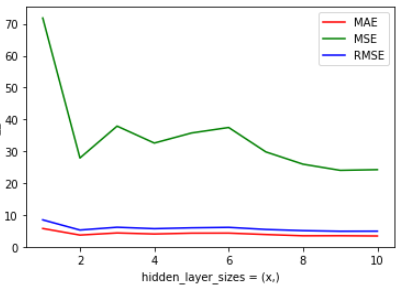
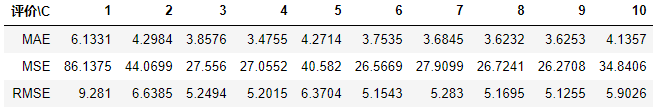
在MAE、MSE和RMSE指标评测下，当第一隐含层有5个神经元，第二个隐含层有7个神经元时效果最好。

1. 第一个隐含层拥有个6神经元



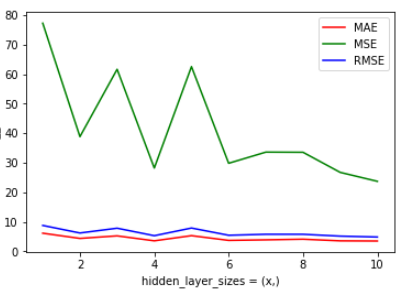
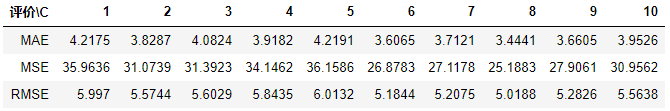
在MAE、MSE和RMSE指标评测下，当第一隐含层有6个神经元，第二个隐含层有3个神经元时效果最好。

1. 第一个隐含层拥有个7神经元



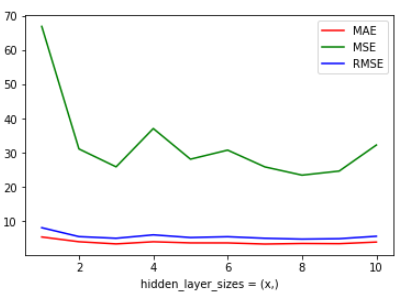
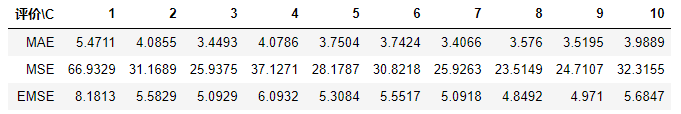
在MAE、MSE和RMSE指标评测下，当第一隐含层有7个神经元，第二个隐含层有2个神经元时效果最好。

1. 第一个隐含层拥有个8神经元



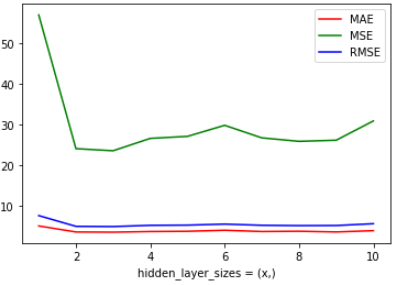
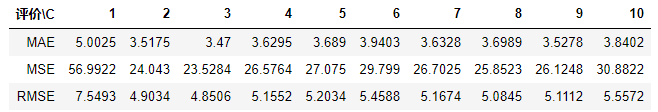
在MAE、MSE和RMSE指标评测下，当第一隐含层有8个神经元，第二个隐含层有4个神经元时效果最好。

1. 第一个隐含层拥有个9神经元



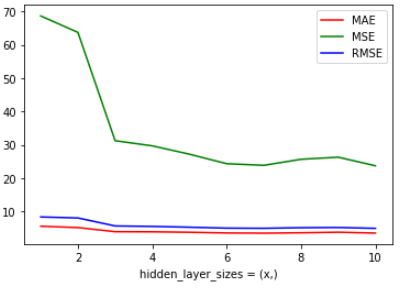
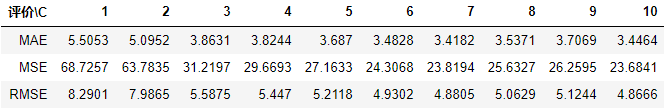
在MAE、MSE和RMSE指标评测下，当第一隐含层有9个神经元，第二个隐含层有3个神经元时效果最好。

1. 第一个隐含层拥有个10神经元



在MAE、MSE和RMSE指标评测下，当第一隐含层有10个神经元，第二个隐含层有2个神经元时效果最好。

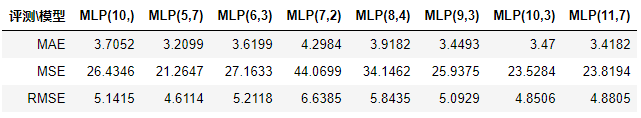
1. 第一个隐含层拥有个11神经元



在MAE、MSE和RMSE指标评测下，当第一隐含层有11个神经元，第二个隐含层有7个神经元时效果最好。

1. **不同网络结构综合对比**

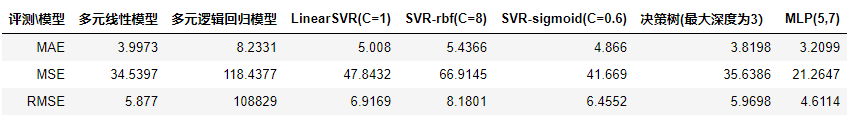
在实验中，第一个隐含层拥有个5~11个神经元，在第一层确定的情况下，第二层的神经元个数从1到10，使用十折交叉验证逐个实验，结果如下表所示：



在MAE、MSE和RMSE指标评测下，对比各个模型，MLP(5,7)性能最优。

1. **不同模型最优对比**

综合不同模型下的最优的评价，得到下表。模型包括多元线性模型、多元逻辑回归模型、LinearSVR(C=1)、SVR-rbf(C=8)、SVR-sigmoid(C=0.6)、决策树(最大深度为3）、MLP(5,7)。模型表现如下：



在MAE、MSE和RMSE指标评测下，多层感知机效果最好，其最优结构为含有两个隐含层，第一个隐含层包含5个神经元，第二个隐含层包含7个神经元及MLP(5,7)。MLP(5,7)在MAE、MSE和RMSE下表现均优于其他模型。