

Requires Changes

**还需满足 1 个要求 变化**

做了很认真的改进！学习态度也很好，希望自己把可选题目完成。但是debug技巧还要多学习～。

**分析数据**

**请求的所有 Boston Housing 数据集统计数据均已得到精确计算。学生可恰当利用 NumPy 功能获得这些结果。**

**学生正确解释各项属性与目标变量增加或减少之间的关联。**

**学生合理解释为何要为某个模型将数据集分解为训练子集和测试子集。训练和测试分解会在代码中正确实现。**

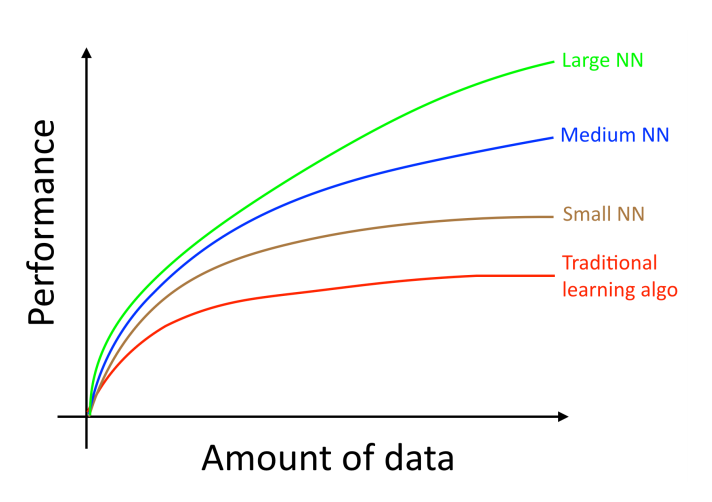
**模型衡量标准**

**性能指标在代码中正确实现。学生正确判断假设模型是否能根据其 R^2 分数成功捕捉目标变量的方差。**

**分析模型的表现**

**随着训练点的不断增加，学生正确判断图表中训练集和验证集曲线的走向并讨论该模型是否会得益于更多的训练点。**

对训练曲线和测试曲线趋势和意义解释的很好。

[](https://udacity-reviews-uploads.s3.amazonaws.com/_attachments/3230/1498724415/ml_curve.png)

传统的机器学习算法（又被称为基于统计的机器学习）在数据量达到一定程度后，更多的数据无法提升模型的表现。深度学习的一个优势就是它可以把大量的数据利用起来，提升学习表现。

**学生提供最大深度为 1 和 10 的分析。如果模型偏差或方差较高，请针对每个图形给出合理的理由。**

**学生根据合理的理由使用模型复杂度图形猜测最优模型的参数。**

**评估模型性能**

**学生准确说明网格搜索算法，并简要探讨该算法的用途。**

**学生准确说明如何对模型进行交叉验证，以及它对网格搜索的作用。**

很好的改进，这里测试集完全不参与。

**学生在代码中正确实现 fit\_model 函数。**

同学，你离成功很近。但是你要自己学会debug。

你的 estimator = {'max\_depth':[1, 2, 3, 4, 5, 6, 7, 8, 9, 10]} 是个字典，让你for循环它的时候，并没有循环它的value。

for i in estimator:

score = cross\_val\_score(i, X, y)

result = (score[0] + score[1] + score[2]) / 3

if result > avl\_r2:

avl\_r2 = result

best\_estimator = DecisionTreeRegressor(max\_depth=i)

你把i打印出来，会发现它不是整数。

当然问题不只这一个，但是这是你要学会的东西！

**学生根据参数调整确定最佳模型，并将此模型的参数与他们猜测的最佳参数进行对比**

**进行预测**

**学生报告表格所列三位客户的预测出售价格，根据已知数据和先前计算出的描述性统计，讨论这些价格是否合理。**

对价格的合理性分析地不错！这里我们可以根据房子的特征来做横向比较；也可以与数据集中特征相近的房子做纵向比较。最终还可以跟统计数据中的最大值，最小值，均值等做比较。

**学生计算了最优模型在测试集上的决定系数，并给出了合理的分析。**

**学生可以合理分析最优模型是否具有健壮性。**

**学生深入讨论支持或反对使用他们的模型预测房屋售价的理由。**

**（可选）预测北京房价**

**学生用代码实现了数据分割与重排、训练模型、对测试集进行测试并返回分数。使用交叉验证对参数进行调优并选出最佳参数，比较两者的差别，最终得出最佳模型对测试集的预测分数。**

**学生的回答与其实现的代码相吻合。并表达了自己的观点。**