



项目

Predicting Boston Housing Prices

此部分属于 Machine Learning Engineer Nanodegree Program

项目审阅

代码审阅

注释

与大家分享你取得的成绩！ 

Meets Specifications

数据研究

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。

numpy是一个很强大的工具，希望你学习这个链接

- <https://docs.scipy.org/doc/numpy/user/quickstart.html>

另外，建议关注无偏估计

- <http://goingmyway.cn/?p=400>

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。

解释地不错，一图胜千言，不过图可以画在这道题这里，而不是在上面画出来

开发模型

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。性能指标在代码中正确实施。

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施。

其实模型训练的本质是为了找到一个训练误差比较小而且对于未知的数据又有比较好的泛化性能。因此，把数据集分成训练集和测试集是为了得到训练误差较低的模型，而测试集是为了检验模型的泛化性能。这里面有一种bias和variance的权衡。

推荐阅读

- <https://www.douban.com/note/325443925/>

分析模型性能

随着训练点的不断增加，学生正确判断图表中训练及测试曲线的走向并讨论该模型是否会得益于更多的训练点。

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。

合理的分析，推荐阅读：<https://www.zhihu.com/question/20448464>

学生根据合理的理由使用模型复杂度图形选择最佳猜测最佳模型。

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。GridSearchCV 会在代码中正确实施。

不错，网格搜索法用于系统地搜索多种参数组合，然后利用交叉验证确定模型的最佳参数组合。

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。除非有合理的理由，否则不得对默认的 3 折交叉验证以外的 GridSearchCV 部分进行修改。

这次理解地很棒！

学生在代码中正确实施 `fit_model` 函数。

```
params = { 'max_depth': [1, 2, 3, 4, 5, 6, 7, 8, 9, 10] }
```

可以这么写

```
params = { 'max_depth': range(1, 11) }
```

学生根据参数调整确定最佳模型，并将此模型与他们选择的模型进行对比。

猜测与结果一致

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。

近邻的思想是不错的想法，不过，建议你多对结果进行分析。

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。

你仿照上面例子的流程做了相似的工作，这样做很好，机器学习建模过程需要这些流程

学生的回答与其实现的代码相吻合。并表达了自己的观点。

思考的不错，这里你可以使用其他的模型进行验证你的想法。

[📄 下载项目](#)