



PROJECT

Predicting Boston Housing Prices

A part of the Machine Learning Engineer Nanodegree Program

PROJECT REVIEW

CODE REVIEW

NOTES

Meets Specifications

同学你好，

你的项目已经非常好了！恭喜你！相信你通过学习和项目已经对这一部分的相关内容有了一个较为全面的了解，这次我只是给你提供了一点建议，让你能够在一些方面再做稍微深入一点的思考。

恭喜你顺利通过！希望你在今后的学习中再接再厉！

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。

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。

开发模型

学生正确判断假设模型是否能根据其 **R²** 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。

很好，你的正确的分析了这个问题。

建议：

你可以查看这个[链接](#)了解一些关于 R^2 的知识。

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。

很棒的代码实现！和全面的分析！

建议：

你可以查看这个[链接](#)以了解更多的知识。

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。

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。

很好，看来你仔细阅读了相关文档，已经对这个有一个比较全面的认识了！

建议：

另外你也可以注意，如果说参数空间非常大的话，我们通常不会直接使用一个完全搜索的方式，这时候就应该考虑使用一些其他方式，比如[RandomizedSearchCV](#)

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。

看来你已经全面理解这个问题了，非常好！

学生在代码中正确实施 `fit_model` 函数。

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。

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。

非常好你的理由已经很充分了！

建议：

请注意这里使用的选择若干个sample，然后对其进行一些深入的分析，从而验证模型的方法是我们在做模型检验中很重要的一种方法，希望你在以后自己的项目中也能够学到这一点。

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。

 [DOWNLOAD PROJECT](#)

[Student FAQ](#)