Predicting Boston Housing Prices

此部分属于 Machine Learning Engineer Nanodegree Program

* [**项目审阅**](https://review.udacity.com/)
* [**代码审阅**](https://review.udacity.com/)
* [**注释**](https://review.udacity.com/)

**SHARE YOUR ACCOMPLISHMENT**

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！

**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。**

你对numpy的使用非常棒！

**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。**

非常好的描述，你清晰的了解了这些特征。

**建议：**

其实你也可以尝试用绘图的方式来佐证你的思考。比如：

import matplotlib.pyplot as plt

plt.figure(figsize=(20, 5))

for i, col in enumerate(features.columns):

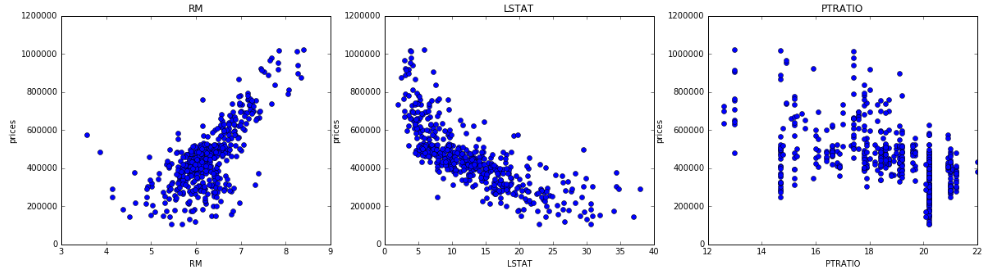
plt.subplot(1, 3, i+1)

plt.plot(data[col], prices, 'o')

plt.title(col)

plt.xlabel(col)

plt.ylabel('prices')

你会得到  
[](https://udacity-reviews-uploads.s3.amazonaws.com/_attachments/31967/1501848846/feature.png)

**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。**

**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。**

不错的回答

**建议：**

你提到“每次选取其中以一个作为测试集”，这里更准确的说法是“验证集”

**学生在代码中正确实现 fit\_model 函数。**

很好的代码实现

**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。**

**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。**

你给出了你对于这些问题的思考，很好！

**建议：**

请注意这些问题是“ 简单地讨论一下你建构的模型能否在现实世界中使用？”的提示而已，所以你还需要做一个总结，得出你的结论。

**（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。**

