

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

 恭喜你通过了这个项目!
后面的学习继续努力呀!

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。



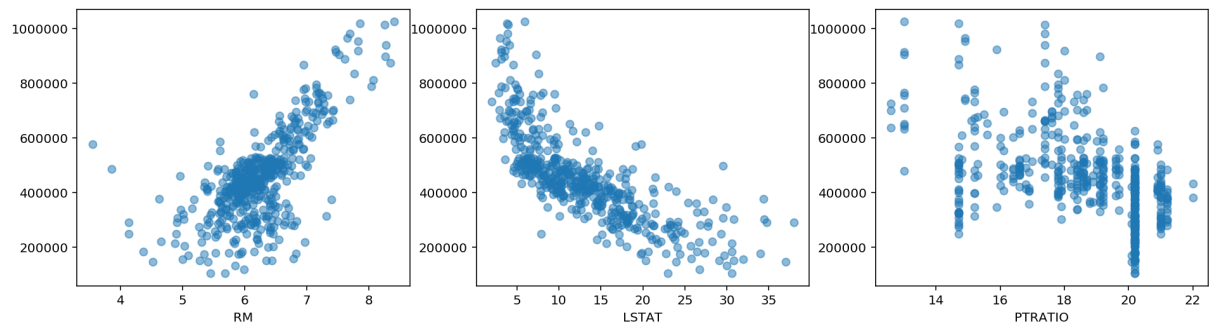
学生正确解释各项属性与目标变量增加或减少之间的关联。

不错的分析，还有一种非常好的思路是通过绘制散点图可视化数据来验证自己的猜想，可以参考下面的代码：

```
# 载入画图所需要的库 matplotlib
import matplotlib.pyplot as plt

# 使输出的图像以更高清的方式显示
%config InlineBackend.figure_format = 'retina'

# 调整图像的宽高
plt.figure(figsize=(16, 4))
for i, key in enumerate(['RM', 'LSTAT', 'PTRATIO']):
    plt.subplot(1, 3, i+1)
    plt.xlabel(key)
    plt.scatter(data[key], data['MEDV'], alpha=0.5)
```



从上图中我们可以看出，虽然‘PTRATIO’与‘MEDV’没有像‘RM’、‘LSTAT’那么强的相关性，但是更倾向于‘PTRATIO’增大‘MEDV’减小。



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。（如果做了可选题，但不正确，算过不通过）



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。

很好的回答。这个[链接](#)可以为你理解欠拟合过拟合提供帮助。这个[知乎讨论](#)简介清晰地介绍了偏差和方差，它所引用的[文章](#)能够帮助你深入理解偏差和方差，都是不错的学习资料。



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。

非常好的补充和修改，很好地解释了K折交叉验证的过程和它对网格搜索的作用。



学生在代码中正确实现 `fit_model` 函数。（如果做了可选题，但不正确，算过不通过）



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。

（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。