**项目概述**

在此项目中，我们将对为马萨诸塞州波士顿地区的房屋价格收集的数据应用基本机器学习概念，以预测新房屋的销售价格。您首先将探索这些数据以获取数据集的重要特征和描述性统计信息。接下来，您要正确地将数据拆分为测试数据集和训练数据集，并确定适用于此问题的性能指标。然后，您将使用不同的参数和训练集大小分析学习算法的性能图表。这让您能够挑选最好地泛化到未见过的数据的最佳模型。最后，您将根据一个新样本测试此最佳模型并将预测的销售价格与您的统计数据进行比较。

**项目亮点**

此项目旨在让您熟练地在 Python 中处理数据集并使用 NumPy 和 Scikit-Learn 应用基本机器学习技术。在使用 sklearn 库中的许多可用算法之前，先练习分析和解释您模型的表现可能有所帮助。

通过完成此项目您将会学习（并最终知道）以下知识：

• 如何使用 NumPy 调查数据集的潜在特征。

• 如何分析各种学习性能图以了解方差和偏差。

• 如何确定用于从未看见的数据进行预测的最佳猜测模型。

• 如何评估模型使用之前的数据处理未看见数据的表现。

**项目描述**

波士顿房屋市场竞争异常激烈，你想成为当地最好的房地产中介。为了与同行竞争，你决定使用几个基本的机器学习概念来帮助你和客户找到其房屋的最佳销售价格。幸运的是，你遇到了波士顿房屋数据集，其中包含大波士顿社区的房屋的各种特征的累积数据，包括其中各个地区的房屋价格的中值。你的任务是利用可用工具基于统计分析来构建一个最佳模型。然后使用该模型估算客户房屋的最佳销售价格。

**软件和库**

对于此项目，你需要安装以下软件和 Python 库：

**•** [**Python 2.7**](https://www.python.org/download/releases/2.7/)

**•** [**NumPy**](http://www.numpy.org/)

**•** [**pandas**](http://pandas.pydata.org/)

**•** [**scikit-learn**](http://scikit-learn.org/)

**•** [**matplotlib**](http://matplotlib.org/)

**•** [**Jupyter Notebook**](http://jupyter.org/)

对jupyter不熟悉的同学可以看一下这两个链接：

**•** [**Jupyter使用视频教程**](http://cn-static.udacity.com/mlnd/how_to_use_jupyter.mp4)

**•** [**为什么使用jupyter？**](https://www.zhihu.com/question/37490497)

如果你还没有安装 Python，我们建议你安装 Python 的发行版 [**Anaconda**](http://continuum.io/downloads) 他包括了我们所有上面提到的工具。安装时选择 Python 2.7 的安装包，**不要**选择 Python 3.x 的。

**开始项目**

要开始这个项目，你可以访问我们的 [**GitHub 页面**](https://github.com/nd009/boston_housing)，或者点击[**这里**](https://github.com/nd009/boston_housing/archive/master.zip)直接下载最新的项目所需文件。

包含三个文件：

• boston\_housing.ipynb: 这是你需要完成项目的主文件。

• housing.csv: 这是项目数据集，你需要加载到 notebook 里。

• visuals.py: 这是帮你创建数据可视化的 python 脚本。

在包含项目文件的命令行或终端下，输入命令 jupyter notebook boston\_housing.ipynb 打开一个浏览器来开始你的项目。根据 notebook 中的项目说明来回答每一个问题。你还可以在 **README** 文件中看到更多项目相关的信息。

**提交项目**

**评审**

优达学城的项目评审师会依据此[**预测波士顿房价项目要求**](https://review.udacity.com/#!/rubrics/268/view)来评审您的提交。提交前请务必确保你已经仔细查看该要求，自己也依此对照检查过了所做的项目。所有要求必须被标注为“合格”，项目才能通过。

**须提交文件**

提交的文件中应包含以下文件并且可以打包为单个 .zip 文件：

• 包含完整实现且可正常运行的代码的 “boston\_housing.ipynb” 文件，并已执行所有代码块和显示了输出。

• 一个由jupyter notebook 导出的 **HTML** 文件，重命名为 **report.html**。此文件需同 ipynb 文件一起提交才能被审阅。

如果你已经核对过了，就可以在下一页提交项目了。

**注意**

**所提交文件的文件名名，包括zip压缩包内的文件名，都不能含有中文及任何ASCII之外的字符，否则会造成提交失败。**

如果你是第一次在优达学城提交项目，点击提交之后，要等1分钟左右才能打开提交页面。如果长时间打不开，可以刷新。如果依然无法打开项目提交页面，可以联系客服微信或者邮件至 support@youdaxue.com。我们正在努力修复这个问题。

项目提交

波士顿房屋市场竞争异常激烈，您想成为当地最好的房地产中介。为了与同行竞争，您决定使用几个基本的机器学习概念来帮助您和客户找到其房屋的最佳销售价格。幸运的是，您遇到了波士顿房屋数据集，其中包含大波士顿社区的房屋的各种特征的累积数据，包括其中各个地区的房屋价格的中值。您的任务是利用可用工具基于统计分析来构建一个最佳模型。然后使用该模型估算客户房屋的最佳销售价格。

**项目文件**

你可以在[**机器学习项目 GitHub**](https://github.com/nd009/boston_housing) 中找到项目文件。请确保使用最新的项目文件完成并提交项目！

**评估**

优达学城的项目评审师会依据此[**预测波士顿房价项目要求**](https://review.udacity.com/#!/rubrics/268/view)来评审您的提交。提交前请务必确保你已经仔细查看该要求，自己也依此对照检查过了所做的项目。所有要求必须被标注为“合格”，项目才能通过。

**提交文件**

提交的文件中应包含以下文件并且可以打包为单个 .zip 文件：

• 包含完整实现且可正常运行的代码的 “boston\_housing.ipynb” 文件，并已执行所有代码块和显示了输出。

• 一个由jupyter notebook 导出的 **HTML** 文件，重命名为 **report.html**。此文件需同 ipynb 文件一起提交才能被审阅。

**提醒：如果你选择上传文档，你需要用英文命名文档，文件名中若有中文字符，会导致提交失败。**

**准备好了！**

当你准备好了提交项目，点击页面下方的项目提交按钮。

**提交**

如果有任何与提交项目相关的疑问，或想查看提交状态，请向我们发送电子邮件 (**support@youdaxue.com**) 或访问我们的[**论坛**](https://discussions.youdaxue.com/)。

**后续步骤**

在项目导师给出反馈后，您会立即收到电子邮件。在此期间，请继续学习后面的项目和内容！

你还没有提交此项目**提交项目**