**项目概述**

欢迎加入“机器学习工程师”纳米学位项目！

在这个项目中，您将创建决策函数，并根据1912年泰坦尼克号海难的乘客特征，如：性别、年龄等，对乘客生还结果进行预测。您可以从一个简单的算法入手，然后逐渐增加该算法的复杂度，直至您至少能精确地预测出所提供数据中80%的乘客的生还结果。通过该项目，您可在正式开始学习本纳米学位前，了解机器学习的一些概念。你还可以在论坛找到该题目在 Kaggle 的[**数据链接**](http://discussions.youdaxue.com/t/project-0-kaggle/7032)。

此外，请确保 Python 装有完成本项目所需的程序包。我们在本项目中将使用到的 Python 库有两个，即 numpy 和 pandas。现在不需担心它们如何运作——我们将在实战项目 1 中接触到它们。本项目还将让您熟悉项目的提交程序，项目提交是您在纳米学位课程中需要完成的内容。

**软件和库**

本项目采用以下软件和 Python 库：

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

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

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

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

你还需要安装和运行 [**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)

如果您还未安装 Python，我们强烈推荐您安装 Python 发行版：[**Anaconda**](http://continuum.io/downloads)，其具备包括上述程序包在内的更多程序包。安装时，确保您选择的是 Python 2.7 安装程序，而不是 Python 3.x 安装程序。

如果您的计算机中已装有 Python 2.7，那么您可使用命令行上的 [**pip**](https://pip.pypa.io/en/stable/) 安装 numpy， scikit-learn 和 Jupyter Notebook（之前叫'iPython'）。如果使用 pip 执行安装时出现问题，[**这个页面**](http://www.lfd.uci.edu/~gohlke/pythonlibs/)对 Windows 用户的某些程序包也是有用的。安装完 pip 之后，你可以执行下列命令安装所需要的包：

sudo pip install numpy pandas matplotlib jupyter scikit-learn

**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/titanic_survival_exploration)，或者点击[**这里**](https://github.com/nd009/titanic_survival_exploration/archive/master.zip)直接下载最新的项目所需文件。

包含三个文件：

• Titanic\_Survival\_Exploration.ipynb: 这是最主要的文件，项目中的主要工作都将在这个文件上完成

• titanic\_data.csv: 项目数据表。您将需要把这个数据加载到 notebook 里。

• titanic\_visualizations.py: 这个 Python 脚本包含 helper 函数，可以让数据和存活结果可视化。

为了打开 jupyter notebook，需要完成以下几步。如果你使用 Windows 系统，你需要打开命令终端或 PowerShell；如果你使用 Mac 或者 Linux 系统，直接打开Terminal 终端即可。使用 cd 命令来打开项目文件夹。例如，在 Windows 上你可以使用 cd C:\Users\username\Documents\ （username 用自己的用户名替换）找到项目所在的文件夹；在 Mac 上，你可以使用 cd ~/Documents/。在 Windows 上你可以使用 dir 命令，在 Mac 或者 Linux 上用 ls 命令列出当前目录中的文件和文件夹。如果发现进错目录，可以使用 cd .. 返回上一级目录。

一旦你进入包含项目文件的文件夹，您可以输入命令

jupyter notebook titanic\_survival\_exploration.ipynb

打开一个浏览器窗口，或者新建标签页，来使用你的 notebook。依照 notebook 上的指导回答每一个问题完成这个项目。我们还提供了随项目的 **READEME** 文档，上面也有关于这个项目的信息和指导。

**项目提交**

**评估**

你的项目会由优达学城项目评审师按照 [**泰坦尼克号探索项目要求**](https://review.udacity.com/#!/rubrics/266/view)进行评审。请确定你仔细阅读了该要求，并在项目提交前自我对检查。要求当中的所有条目都必须**合格**项目才能通过。

**提交文件**

当你准备好提交项目时，你可以把所有文件压缩成一个 zip 压缩包上传，或者放在你的 GitHub Repo 中的文件夹 titanic\_survival\_exploration下，并通过 Github 提交。

需要提交的文件有：

• titanic\_survival\_exploration.ipynb: 带有完整问题答案和代码的 notebook 文件。

• report.html: notebook 项目导出的 **HTML** 文件。

**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。**

当你准备好所有这些文件，并且依照项目要求核对过之后，就可以在下面的项目提交页面提交你的项目了。

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（ 我们正在努力修复这个问题）。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