

# 第一周

---

## 工具安装

---

在这第一个部分，大家要完成两个工具的安装，分别是 Anaconda 和 Jupyter notebook：

- Anaconda 是一个包含数据科学常用包的发行版本。它基于 conda —— 一个包和环境管理器 —— 衍生而来。你将使用 conda 创建环境，以便分隔使用不同 Python 版本和/或不同包的项目。你还将使用它在环境中安装、卸载和更新包。通过使用 Anaconda，使我处理数据的过程更加愉快。
  - 下载地址：<https://www.continuum.io/downloads>（推荐使用 2.7 版本）
  - 安装教程：[Windows](#), [Linux](#), [Mac](#)
  - 安装完成后输入 `conda install numpy scipy pandas scikit-learn` 命令安装所需要的包
- Jupyter notebook 是 Web 文档，能让你将文本、图像和代码全部组合到一个文档中。它已经成为数据分析的标准环境。
  - Anaconda 中包含 Jupyter notebook，因此可以直接按教程开始使用，如果不行，可使用 `conda install jupyter notebook` 命令安装
  - [快速入门教程](#)
  - [视频教程\(中文\)](#)

之后还需要初步掌握 Python 在机器学习、数据分析中最常用的两个库 Pandas 和 Numpy:

- [\(译\) 十分钟的 pandas 入门教程](#)
- [\(译\) NumPy教程\(一\)](#)
- 注：在命令行中输入 `ipython` 即可进入 Python 环境

## 内容

---

- Python 编程基础
  - [编程基础：Python — 学习面向对象编程方法](#)
  - [数据分析入门 — 使用 NumPy 和 Pandas 进行数据分析](#)

- 其他前置学习基础
  - [Windows 命令行基础](#) 了解 1,2,3 即可
  - [Linux\(Mac\) 命令行基础](#)
- 图像表示
  - 图像特征
    - [颜色直方图](#)
    - [目标检测的图像特征提取之（一）HOG特征](#)
    - [目标检测的图像特征提取之（二）LBP特征](#)
    - [目标检测的图像特征提取之（三）Haar特征](#)
- **VIDEO**, Machine Learning (2017, Spring), 李宏毅, 台湾大学
  - [Regression](#)
  - [Where does the error come from?](#)
  - [Gradient Descent](#)
  - 视频也可以用 FTP 在 10.65.1.69 服务器上获取, 速度更快
- 其他
  - [各种距离](#)
  - [机器学习中的Bias\(偏差\), Error\(误差\), 和Variance\(方差\)有什么区别和联系?](#)
  - [Why Data Scientists Split Data into Train and Test](#)
  - [K折交叉验证评估模型性能](#)
  - 一个介绍机器学习比较通俗化的系列文章
    - [写给大家看的机器学习书 第一篇](#)
    - [写给大家看的机器学习书 第二篇](#)
    - [写给大家看的机器学习书 第三篇](#)
    - [写给大家看的机器学习书 第四篇](#)

## 作业

---

两个作业均使用 Jupyter Notebook, 完成其中的 TO DO 代码和 question (问题) 后, 需提交以下文件:

- 包含完整实现且可正常运行的代码的 “boston\_housing.ipynb” 和 “features.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, 分别重命名为 boston\_report.html 和 features\_report.html。这两个文件需同 ipynb 文件一起提交
- 完成作业后请将上述文件压缩为 .zip 文件后作为附件发送至

## 图像特征提取练习

自己编写代码提取图片的 HOG 特征。（作业文件中 **k-NN on features** 部分暂时不用完成）

作业文件：assignment/Image\_features\_exercise/features.ipynb

（本作业修改自 [CS231n assignment 1](#)）

## 波士顿房价预测项目

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本周学到的几个机器学习概念，以预测新房屋的销售价格。你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，你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，你将自己编写一个线性回归的模型，并使用不同的参数和训练集大小分析学习算法的性能图表。最后，你将根据一个新样本测试此模型并将预测的销售价格与你的统计数据进行比较。

作业文件：assignment/boston\_housing/boston\_housing.ipynb

（本项目作业修改自 [Udacity Machine Learning Nanodegree Project](#)）

## 其他

---

如果大家在课程学习中或者做作业时遇到问题，尽量使用 Twist 进行沟通，暑期班内容的变化动态 Twist 也能自动获取。