**Model Evaluation and Validation**

**Project: Predicting Boston Housing Prices**

**Install**

This project requires **Python** and the following Python libraries installed:

* [NumPy](http://www.numpy.org/)
* [Pandas](http://pandas.pydata.org/)
* [matplotlib](http://matplotlib.org/)
* [scikit-learn](http://scikit-learn.org/stable/)

You will also need to have software installed to run and execute a [Jupyter Notebook](http://ipython.org/notebook.html)

If you do not have Python installed yet, it is highly recommended that you install the [Anaconda](http://continuum.io/downloads) distribution of Python, which already has the above packages and more included.

**Code**

Template code is provided in the boston\_housing.ipynb notebook file. You will also be required to use the included visuals.py Python file and the housing.csv dataset file to complete your work. While some code has already been implemented to get you started, you will need to implement additional functionality when requested to successfully complete the project. Note that the code included in visuals.py is meant to be used out-of-the-box and not intended for students to manipulate. If you are interested in how the visualizations are created in the notebook, please feel free to explore this Python file.

**Run**

In a terminal or command window, navigate to the top-level project directory boston\_housing/ (that contains this README) and run one of the following commands:

ipython notebook boston\_housing.ipynb

or

jupyter notebook boston\_housing.ipynb

This will open the Jupyter Notebook software and project file in your browser.

**Data**

The modified Boston housing dataset consists of 489 data points, with each datapoint having 3 features. This dataset is a modified version of the Boston Housing dataset found on the [UCI Machine Learning Repository](https://archive.ics.uci.edu/ml/datasets/Housing).

**Features**

1. RM: average number of rooms per dwelling
2. LSTAT: percentage of population considered lower status
3. PTRATIO: pupil-teacher ratio by town

**Target Variable** 4. MEDV: median value of owner-occupied homes

**Model Evaluation and Validation**

**Project: Predicting Boston Housing Prices**

**Project Overview**

The Boston housing market is highly competitive, and you want to be the best real estate agent in the area. To compete with your peers, you decide to leverage a few basic machine learning concepts to assist you and a client with finding the best selling price for their home. Luckily, you’ve come across the Boston Housing dataset which contains aggregated data on various features for houses in Greater Boston communities, including the median value of homes for each of those areas. Your task is to build an optimal model based on a statistical analysis with the tools available. This model will then be used to estimate the best selling price for your clients' homes.

**Project Highlights**

This project is designed to get you acquainted with the many techniques for training, testing, evaluating, and optimizing models, available in sklearn.

Things you will learn by completing this project:

* How to explore data and observe features.
* How to train and test models.
* How to identify potential problems, such as errors due to bias or variance.
* How to apply techniques to improve the model, such as cross-validation and grid search.

**Starting the Project**

For this assignment, you can find the boston\_housing folder containing the necessary project files on the [**Machine Learning projects GitHub**](https://github.com/udacity/machine-learning), under the projects folder. You may download all of the files for projects we'll use in this Nanodegree program directly from this repo. Please make sure that you use the most recent version of project files when completing a project!

This project contains three files:

* boston\_housing.ipynb: This is the main file where you will be performing your work on the project.
* housing.csv: The project dataset. You'll load this data in the notebook.
* visuals.py: This Python script provides supplementary visualizations for the project. Do not modify.

In the Terminal or Command Prompt, navigate to the folder containing the project files, and then use the command jupyter notebook boston\_housing.ipynb to open up a browser window or tab to work with your notebook. Alternatively, you can use the command jupyter notebook or ipython notebook and navigate to the notebook file in the browser window that opens. Follow the instructions in the notebook and answer each question presented to successfully complete the project. A **README** file has also been provided with the project files which may contain additional necessary information or instruction for the project.

**Submitting the Project**

**Evaluation**

Your project will be reviewed by a Udacity reviewer against the [**Predicting Boston Housing Prices project rubric**](https://review.udacity.com/#!/rubrics/103/view). Be sure to review this rubric thoroughly and self-evaluate your project before submission. All criteria found in the rubric must be *meeting specifications* for you to pass.

**Submission Files**

When you are ready to submit your project, collect the following files and compress them into a single archive for upload. Alternatively, you may supply the following files on your GitHub Repo in a folder named boston\_housing for ease of access:

* The boston\_housing.ipynb notebook file with all questions answered and all code cells executed and displaying output.
* An **HTML** export of the project notebook with the name **report.html**. This file *must* be present for your project to be evaluated.

**Using your workspace (in the next node) you can submit your project directly for review. Remember to upload the HTML report once you are done working on the Jupyter notebook into your workspace BEFORE submitting your report.**

Once you have collected these files and reviewed the project rubric, proceed to the project submission page.

**Project Submission**

The Boston housing market is highly competitive, and you want to be the best real estate agent in the area. To compete with your peers, you decide to leverage a few basic machine learning concepts to assist you and a client with finding the best selling price for their home. Luckily, you’ve come across the Boston Housing dataset which contains aggregated data on various features for houses in Greater Boston communities, including the median value of homes for each of those areas. Your task is to build an optimal model based on a statistical analysis with the tools available. This model will then be used to estimate the best selling price for your clients' homes.

**Project Files**

For this assignment, you can find the boston\_housing folder containing the necessary project files on the [**Machine Learning projects GitHub**](https://github.com/udacity/machine-learning), under the projects folder. You may download all of the files for projects we'll use in this Nanodegree program directly from this repo. Please make sure that you use the most recent version of project files when completing a project!

**Evaluation**

Your project will be reviewed by a Udacity reviewer against the [**Predicting Boston Housing Prices project rubric**](https://review.udacity.com/#!/rubrics/103/view). Be sure to review this rubric thoroughly and self-evaluate your project before submission. All criteria found in the rubric must be *meeting specifications* for you to pass.

**Submission Files**

When you are ready to submit your project, collect the following files and compress them into a single archive for upload. Alternatively, you may supply the following files on your GitHub Repo in a folder named boston\_housing for ease of access:

* The boston\_housing.ipynb notebook file with all questions answered and all code cells executed and displaying output.
* An **HTML** export of the project notebook with the name **report.html**. This file *must* be present for your project to be evaluated.

**I'm Ready!**

When you're ready to submit your project, click on the **Submit Project** button at the bottom of the page.

If you are having any problems submitting your project or wish to check on the status of your submission, please email us at **machine-support@udacity.com** or visit us in the [**discussion forums**](http://discussions.udacity.com/).

**What's Next?**

You will get an email as soon as your reviewer has feedback for you. In the meantime, review your next project and feel free to get started on it or the courses supporting it!