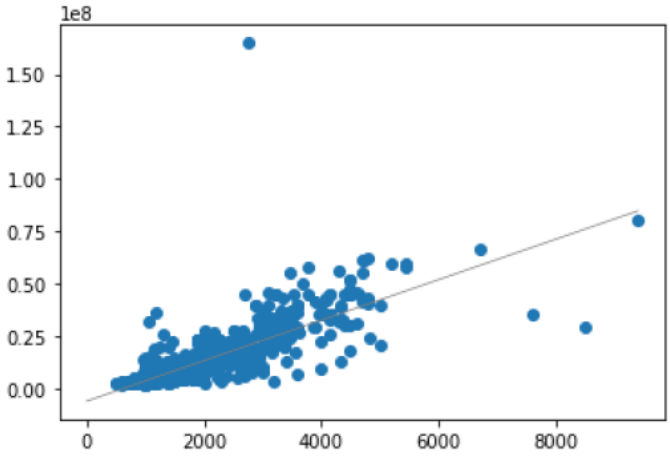
COMP-2704:
Supervised Machine Learning



Assignment 1: Linear Regression

Setup

First, complete the following steps to setup this assignment.

i. Open Jupyter Notebook in a web browser.
ii. Create a folder named "SupervisedML" and navigate into this folder.
iii. Within the "SupervisedML" folder, create a folder named "Assignment1" and navigate into this folder. The path to this folder should be "~/SupervisedML/Assignment1".
iv. Open the link: https://github.com/luisguiserrano/manning/tree/master/Chapter_3_Linear_Regression
v. Download the files:
    - *House_price_predictions.ipynb*
    - *utils.py*
    - *Hyderabad.csv*
vi. Upload these files to the folder you created "~/SupervisedML/Assignment1".
vii. Open the *House_price_predictions.ipynb* notebook and run all code cells. Fix any errors that occur.

Problem

To complete the first problem, open and run all cells in the *House_price_predictions.ipynb* notebook.

1) *[2 marks]* Using *model*, predict the price of a house with three bedrooms and an area of 1000 square feet. Provide an uncertainty for this prediction.

To complete the remainder of the problems, create a notebook with filename *SML_a1.ipynb* within the folder "~/SupervisedML/Assignment1". Write code to complete the steps below and add markdown cells to answer questions. You may copy relevant lines of code from *House_price_predictions.ipynb*.

First, import the necessary libraries and modules. Then Load the data from *Hyderabad.csv* into an SFrame named "data". Now, do the following:

2) *[2 marks]* Show two scatter plots: *Price* vs. *Area* and *Price* vs. *No. of Bedrooms*.

3) *[4 marks]* Create a model called *two_feature_model* that uses *Price* as the target, and *Area* and *No. of Bedrooms* as features. Train the model and list the coefficients of the optimal solution.

4) *[2 marks]* What is the maximum error and root-mean-squared error of *two_feature_model*? Compare these values with the errors for *model* and state which is better.

5) *[4 marks]* Use your trained model to predict the price of the following houses:
   - 4000 square feet and 5 bedrooms,
   - 1500 square feet and 3 bedrooms,
   - 1000 square feet and 2 bedrooms.

   Provide an uncertainty for your predictions.

6) *[2 marks]* With the exception of import and print statements, add a comment before each line of code in *SML_a1.ipynb* to explain what it does.

Contributions

Using markdown at the end of the notebook, list the contribution of each student to the assignment, referencing specific question numbers and other tasks such as formatting and submitting. Ideally, both students will contribute to all questions.

Submission

Upload your two notebooks to the Assignment 1 dropbox on the course website. Late submissions will lose 10%.

Total marks = 16