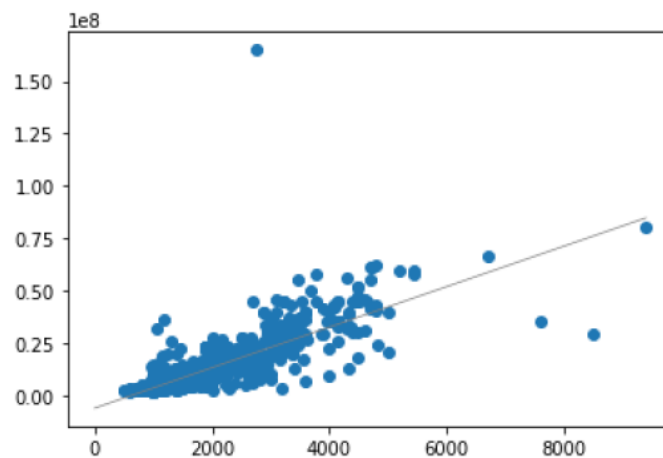


COMP-2704:

Supervised Machine Learning



Assignment 1: Linear Regression

You are to work collaboratively with your group members to complete this cooperative learning assignment. Choose among your group who will be:

- Facilitator – keeps group working toward completion and schedules after class meetings.
- Evaluator – ensures all questions have been answered and documents are formatted neatly; makes final submission to dropbox.

Indicate at the top of your document who has been selected for each role. Note that Assignment 2 will be done by the same groups, and different students will need to take on the facilitator and evaluator roles for that assignment.

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

- 1) Use the *House_price_predictions.ipynb* notebook to answer the following questions. Insert text cells to write out your answers; be sure to state the question number.
 - a) [2 marks] How many rows and columns are in the data file?
 - b) [2 marks] Just from looking at the data, are there any potential outliers shown in the price vs. area scatter plot? List the (x, y) coordinates of any potential outliers.
 - c) [2 marks] What is the equation of the best fit line produced by *simple_model* that uses only price and area?
 - d) [2 marks] What coefficient values are associated with the intercept and area in the trained *model* that uses all features? Are these values the same as the intercept and slope found by *simple_model*?
 - e) [3 marks] What is the maximum error and root-mean-squared error of *model*? Explain what these mean.
 - f) [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.
- 2) Create a notebook with filename *SML_a1_q2.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* (but notice step 3 below).

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

- a) [2 marks] Show two scatter plots: *Price vs. Area* and *Price vs. No. of Bedrooms*.
 - b) [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.
 - c) [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.
 - d) [4 marks] Use your trained model to predict the price, along with uncertainties, of the following houses:
 - 4000 square feet and 5 bedrooms,
 - 1500 square feet and 3 bedrooms,
 - 1000 square feet and 2 bedrooms.
- 3) [10 marks] With the exception of import and print statements, add a comment before each line of code in *SML_a1_q2.ipynb* to explain what it does.

Peer Evaluation

[3 marks] Each group member is to review the notebooks that were created for the assignment. Then, fill in the table below using an integer from 0 to 3 to indicate how well each item was done. Accurately filling in this table is worth 3 marks.

Student:	Name 1	Name 2	Name 3	Name 4
Problem step 1)				
Problem step 2)				
Problem step 3)				
Every member contributed				
Every member treated with respect				

Submission

Upload your two notebooks to the Assignment 1 dropbox on the course website. Copy the peer evaluation table into the dropbox comment window. Late submissions will lose 10%.

Total marks = 38