Problem Statement - Part I

This assignment contains two parts. Part-I is a programming assignment (to be submitted in a Jupyter Notebook), and Part-II includes subjective questions (to be submitted in a PDF file).

Part-II is given on the next page.

Assignment Part-I

A US-based housing company named Surprise Housing has decided to enter the Australian market. The company uses data analytics to purchase houses at a price below their actual values and flip them on at a higher price. For the same purpose, the company has collected a data set from the sale of houses in Australia. The data is provided in the CSV file below.

The company is looking at prospective properties to buy to enter the market. You are required to build a regression model using regularisation in order to predict the actual value of the prospective properties and decide whether to invest in them or not.

The company wants to know:

- Which variables are significant in predicting the price of a house, and

- How well those variables describe the price of a house.

Also, determine the optimal value of lambda for ridge and lasso regression.

Business Goal

You are required to model the price of houses with the available independent variables. This model will then be used by the management to understand how exactly the prices vary with the variables. They can accordingly manipulate the strategy of the firm and concentrate on areas that will yield high returns. Further, the model will be a good way for management to understand the pricing dynamics of a new market.

Downloads

You can download the data set file from the link given below:

📄 **Assignment - Dataset**	⬇ **Download**

Data Definition

The details of the various variables are given in the file below.

📄 **Data Definition**	⬇ **Download**

The next page contains the problem statement of Part-II of the assignment.

Report an error

Problem Statement - Part II

Assignment Part-II

The following questions are the second part of the graded assignment. Please submit the answers in one PDF file. For writing normal text, please use MS Word (or similar software that can convert documents to PDF). For equations and figures, you can write/draw them on a blank sheet of paper using a pen, click images and upload them in the same Word document.

The final submission will be in the form of one PDF file. A sample PDF to illustrate the submission format is provided below.

Note: **DO NOT** copy or paste answers from *anywhere,* and type the answers in your own words; your solution files will be tested using automatic plagiarism checkers and will attract a heavy penalty if plagiarism is detected.

Please limit your answers to less than 500 words per question.

Question 1

What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

Question 2

You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

Question 3

After building the model, you realised that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?

Question 4

How can you make sure that a model is robust and generalisable? What are the implications of the same for the accuracy of the model and why?

The sample solution format is attached below.

📄　**Subjective Questions - Sample PDF Format**　　　⬇ Download

The evaluation rubrics for both parts of the assignment are given on the next page.

❗ <u>Report an error</u>