



Bootcamp AI / Machine Learning

Multiple Linear Regression

Summary: In this Module, you will learn about Multiple Linear Regression.

Version: 1.00

Contents

I	Introduction	2
II	General instructions	3
III	Exercise 00	4
IV	Exercise 01	5
V	Exercise 02	6
VI	Exercise 03	8
VII	Exercise 04	9
VIII	Bonus part	10
IX	Submission and peer-evaluation	11

Chapter I

Introduction

Greetings!



If you haven't already done so, read `en.toolkit.pdf`.

What this Module will cover:

In this module, we will delve into the practical application of multiple linear regression, building on the foundational concepts of simple linear regression. Through hands-on exercises, we'll gain insight into how multiple independent variables can be utilized to predict a dependent variable. By using essential libraries and tools like pandas for data manipulation and matplotlib or seaborn for visualization, we'll explore how to implement multiple linear regression models effectively. While the principles remain rooted in simple linear regression, the extension to multiple variables adds complexity and richness to the predictive modeling process. By the end of this module, you'll be equipped with the skills necessary to tackle real-world datasets using multiple linear regression techniques, solidifying your understanding of predictive analytics and data-driven decision-making.

Best of luck to all.

Chapter II


General instructions

Unless explicitly specified, the following rules will apply every day of this Bootcamp.

- This subject is the one and only trustable source. Don't trust any rumor.
- This subject can be updated up to one hour before the turn-in deadline.
- The assignments in a subject must be done in the given order. Later assignments won't be rated unless all the previous ones are perfectly executed.
- Be careful about the access rights of your files and folders.
- Your assignments will be evaluated by your peers.
- You must not leave in your turn-in your workspace any file other than the ones explicitly requested By the assignments.
- You have a question? Ask your left neighbor. Otherwise, try your luck with your right neighbor.
- Every technical answer you might need is available in the **man** or on the Internet.
- By Thor, by Odin! Use your brain!!!

Chapter III


Exercise 00

	Exercise 00
Data exploration	
Turn-in directory : <i>ex00/</i>	
Files to turn in : Beginner02.ipynb	
Allowed functions : A11	

For this first exercise, you'll need to load the data into colab and do some data exploration, i.e. you'll need to understand your data because this time you don't have just two collones as in the previous modules.

Chapter IV


Exercise 01

	Exercise 01
Data Preparation	
Turn-in directory : <i>ex01/</i>	
Files to turn in : Beginner02.ipynb	
Allowed functions : A11	

For this second exercise, we're going to clean up the data so that we can make better use of it.

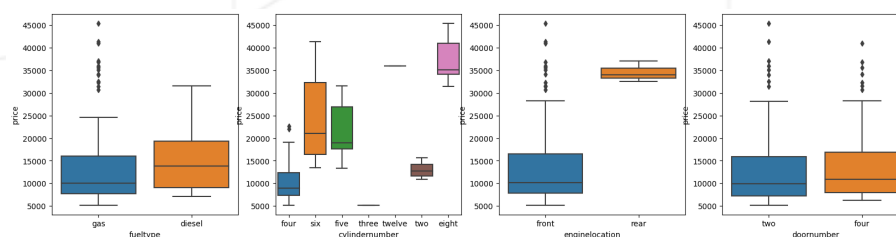
Chapter V

Exercise 02

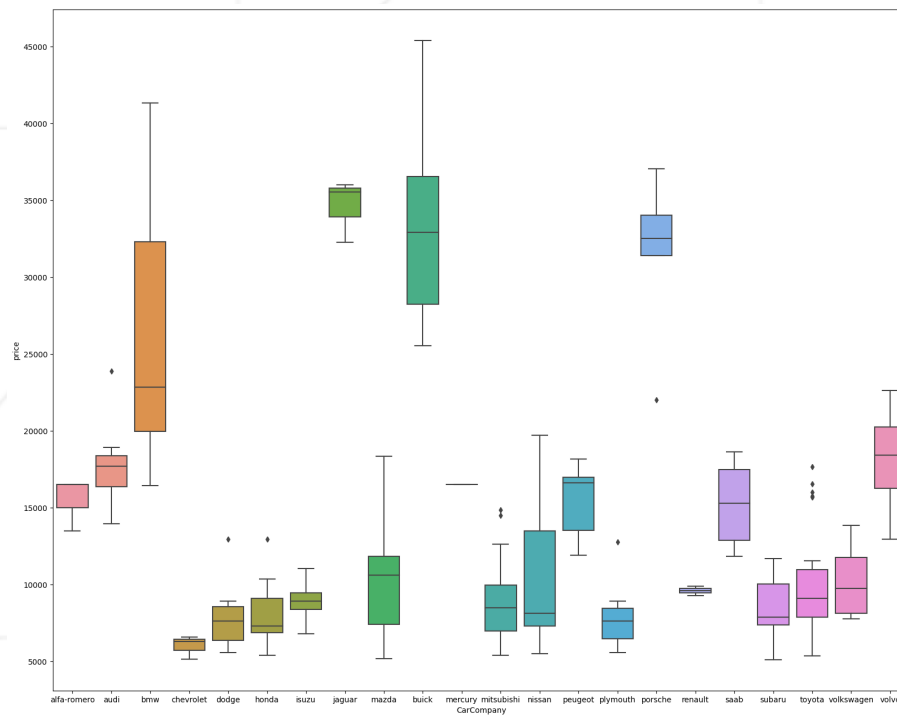
	Exercise 02
Data Visualisation	
Turn-in directory : <i>ex02/</i>	
Files to turn in : Beginner02.ipynb	
Allowed functions : All	

In this exercise, you will explore and visualize the relationship between categorical variables and car prices using box plots. The dataset you will work with is the "CarData" dataset, which contains various attributes related to cars, including categorical features such as fuel type, cylinder number, engine location, and door number. By creating box plots for these categorical variables against the car prices, you can gain insights into how different categorical attributes impact the pricing of cars.

You should have something like this:




You should have something like this:




Chapter VI

Exercise 03

	Exercise 03
Data Augmentation	
Turn-in directory : <i>ex03/</i>	
Files to turn in : Beginner02.ipynb	
Allowed functions : All	

In this exercise, you will explore the concepts of data augmentation and feature engineering in the context of car pricing analysis. The dataset you will work with contains information about various car attributes, including the car company, price, and other categorical variables. You will perform data augmentation by dividing the car prices into buckets based on predefined ranges and create new features using one-hot encoding to enhance the dataset for modeling purposes.


id	color
1	red
2	blue
3	green
4	blue



id	color_red	color_blue	color_green
1	1	0	0
2	0	1	0
3	0	0	1
4	0	1	0

Chapter VII

Exercise 04

	Exercise 04
Model	
Turn-in directory : <i>ex04/</i>	
Files to turn in : Beginner02.ipynb	
Allowed functions : A11	

In this exercise, you will walk through the process of building a linear regression model for predicting car prices using a dataset containing various car attributes. You will perform the necessary preprocessing steps, including splitting the data into training and testing sets, applying Min-Max scaling, and finally, training and evaluating the linear regression model.

With this function you should get results higher than 0.8:

```
print(r2_score(y_true=y_test, y_pred=y_pred))
```

Chapter VIII

Bonus part

If you have some time, you can try to validate this little bonus. It is not mandatory, so don't waste too much time doing it.

You'll have to redo this whole module without the libraries, you can start with the simple parts like the r^2 score, look for the mathematical formulas on the internet.



The bonus part will only be assessed if the mandatory part is PERFECT. Perfect means the mandatory part has been integrally done and works without malfunctioning. If you have not passed ALL the mandatory requirements, your bonus part will not be evaluated at all.

Chapter IX

Submission and peer-evaluation

- Create a `professional_training_beginner` folder at the root of your home, and move around in it.
- Create a new `module02` folder and navigate to it.



Please note, during your defense anything that is not present in the folder for the day will not be checked.