IBB31103 | ISB46703 Introduction to/Principles of Artificial Intelligence

Universiti Kuala Lumpur
February 2023
Dr. Faiz

Assignment

Weight: 10%
Due: 23 June 2023

1 Instructions

In this assignment, you will explore the concept of multiple linear regression and the importance of linearity in regression analysis using a provided stock dataset. You will also learn how to visualise linearity using scatter plots and evaluate the assumptions of multiple linear regression.

2 Dataset

You are provided with a dataset containing the following variables:

Variable	Description
`Year`	Year of stock observation
`Month`	Month of stock observation
`Interest Rate`	Stock interest rate
`Unemployment Rate`	Number of unemployed people divided by total number in the labour force, multiplied by 100
`Stock Index Price`	Total share prices included in the index

Table 1: Stock dataset.

Dataset download: link:

3 Tasks

3.1 Dataset (10 points)

1. Download dataset.
2. Decompress dataset.
3. Load the dataset into `pandas` dataframe.
4. Display dataframe summary.

3.2 Linearity Visualisation (15 points)

1. Explain the importance of linearity in multiple linear regression.
2. Define linearity and its implications for the relationship between the independent and dependent variables.
3. Choose two variables from the dataset in 3.1 and create a scatter plot to visualise their relationship. Include proper labeling and formatting in the plot.
4. Based on the scatter plot, discuss whether the relationship between the variables appears to be linear or non-linear. Provide reasoning to support your conclusion.

3.3 Multiple Linear Regression (20 points)

1. Define multiple linear regression and explain its purpose in statistical analysis.
2. Discuss the assumptions of multiple linear regression and explain why they are important.
3. Provide a step-by-step explanation of the process to perform multiple linear regression analysis, including data preparation, model building, coefficient estimation, and model evaluation.
4. Use the provided dataset and perform a multiple linear regression analysis. Include the following steps:
    - Use `Interest Rate` and `Unemployment Rate` as independent variables and `Stock Index Price` as dependent variable.
    - Build a multiple linear regression model using `sklearn.linear_model.LinearRegression()`.
    - Interpret the coefficients of the model and assess their significance.
    - Evaluate the model's root mean squared error (RMSE) and interpret the results.
    - Predict `Stock Index Price` when `Interest Rate` = 3.7 and `Unemployment Rate` = 1.2
    - Discuss any limitations or potential issues with the model.

4 Submission

Submit the following materials to VLE by **23 June 2023 (Friday) 5:00pm** :

- Fully/clearly commented Jupyter Notebook (`.ipynb`) or code files used of the regression analysis and via VLE. Use Markdown cells when answering essay questions.

5 Statement on collaboration and individual submission, and plagiarism policy

While collaboration is permitted, it is strictly prohibited to engage in any form of plagiarism or copying. Plagiarism involves using someone else's work, ideas, or words without proper acknowledgment. It is a serious academic offense and goes against the principles of academic integrity. Any instances of plagiarism or copying will be thoroughly investigated, and appropriate penalties will be applied as per UniKL MIIT policies. This may include receiving a failing grade on the assignment or facing further disciplinary actions.

5.1 Avoiding plagiarism

Please ensure that all the work you submit is entirely your own. Any sources used in your assignment should be properly cited and referenced. This includes both direct quotes and paraphrased information.