Multiple Regression Analysis Exercises: (Deadline: $JUNE,\ 6^{th}$)

1. Import the necessary libraries: numpy, pandas, matplotlib.pyplot, and sklearn.
2. Load a dataset that contains multiple variables (features) and a target variable (dependent variable). You can use a publicly available dataset or one of your choice.
3. Perform data preprocessing tasks such as handling missing values, encoding categorical variables, and scaling the features if necessary.
4. Split the dataset into training and testing sets using a 80:20 ratio.
5. Fit a multiple regression model to the training data.
6. Evaluate the model's performance on the testing data using appropriate evaluation metrics such as mean squared error (MSE) or R-squared.
7. Perform feature selection or feature engineering techniques to improve the model's performance if needed.
8. Visualize the predicted values versus the actual values using a scatter plot or other appropriate plots.
9. Interpret the model coefficients to understand the impact of each feature on the target variable.
10. Try different variations of the model by adding interaction terms or applying regularization techniques such as Lasso or Ridge regression.
11. Compare the performance of the different models and select the best one based on the evaluation metrics.
12. Finally, make predictions on new, unseen data using the selected model.