Graduate Certificate in Artificial Intelligence with Machine Learning
AIGC 5002 - Machine Learning and Deep Learning
Fall 2023

Lab 2:
Multiple Linear Regression (MLR)
September 20, 2023

Submission guidelines:

- For this lab, you will need to submit 1 PDF file by the end of lab time.
- After you complete all the exercises, convert your Jupyter Notebook (.ipynb) to PDF. Name the PDF as follows: firstname_lastname_LAB2.pdf
- Go to the course Blackboard → Labs folder → Lab Exercise 2 and submit the pdf.

Part 1: MLR

Assuming we're working with a software company that's trying to predict a product's user satisfaction score. This score is dependent on four factors:

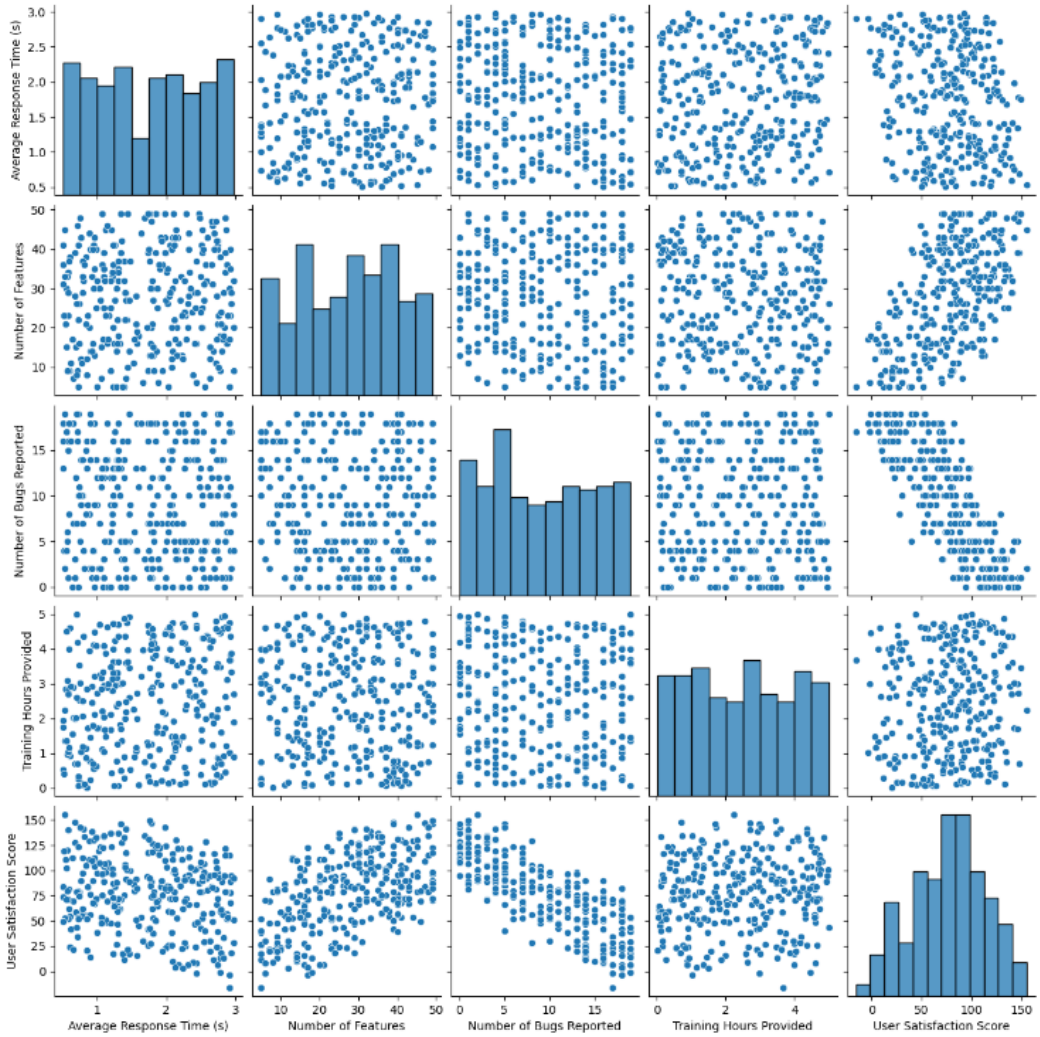
- Average response time (in seconds) of the application.
- Number of features in the product.
- Number of bugs reported by users.
- Training hours provided to users.

We're given a dataset collected by the company so we can use it to create an (MLR) model.

1. Download the dataset (.csv) from the BlackBoard → Lab 2
2. Read the dataset into a DataFrame.
3. Explore the dataset by plotting all independent variables versus the dependent variable (example below). (Hint:

```
import seaborn as sns
sns.pairplot(df)
plt.show()
```

)



4. Clearly state your observations using a markdown cell.
5. Split your dataset to training and testing. (State your choice of split method and ratios and justify your choices)
6. Fit an MLR model to the data.
7. Compute the necessary evaluation metrics, justify your choice, and analyze the results (was the model able to successfully predict and generalize)
8. Demo your code.

Part 2: MLR vs LR

(Note: LR and MLR using LinearRegression() are computed using closed-form methods (A closed-form solution is an equation that can be solved in terms of functions and mathematical operations.). You can read more on how to design these models using an iterative method such as Stochastic Gradient Decent (SGD) through SDGRegressor() in Sklearn. In iterative methods, you can control parameters such as the learning rate, number of iterations, and tolerance.)

1. Choose one independent variable from the dataset and design a simple LR model.
2. Compare your results and state your analysis.

Enjoy!
