

COMP 4983: Lab Exercise #2

Mark: /35

[Due: Sep 16, 2022 @1730
Assignment Submission
Folders]

Instructions:

In this lab, you will use Python to

- create a machine learning model that predicts the price of an automobile
- find the feature that yields the lowest mean absolute error (MAE)

Part 1: Linear Regression using Python

[20 marks] In Lab Exercise #1, you used AML to build a linear regression model to predict the price of an automobile. In this part of the lab, you will perform the same task, but using Python and additional packages, including

- random
(<https://docs.python.org/3.7/library/random.html>)
- pandas
(<https://pandas.pydata.org/pandas-docs/stable/>)
- NumPy
(<https://docs.scipy.org/doc/numpy/user/quickstart.html>)
- scikit-learn
(http://scikit-learn.org/stable/tutorial/statistical_inference/supervised_learning.html#linear-regression)

1. Download the raw automobile price data, *AutomobilePrice_Lab2.csv*, from BCIT Learning Hub (Content | Laboratory Material | Lab 2) and save it in your working directory.
2. Download *AutomobilePricePred.py* from BCIT Learning Hub (Content | Laboratory Material | Lab 2) and save it as *AutomobilePricePred_Lab2.py* in your working directory. This script contains the skeleton for creating, training and testing a linear regression model for the automobile price prediction. Your task is to modify *AutomobilePricePred_Lab2.py* by replacing ??? with appropriate code.
3. Run *AutomobilePricePred_Lab2.py*.

Your results should be comparable (dependent on the training data and test data used) to the following evaluation result metrics:

| | |
|------------------------------|-------------|
| Mean Absolute Error | 1656.147651 |
| Root Mean Squared Error | 2456.983209 |
| Coefficient of Determination | 0.910392 |

Part 2: Linear Regression Model Feature Selection

[15 marks] In this part of the lab, you will use the linear regression model implementation from Part (1), i.e., `LinearRegression.fit()` and `LinearRegression.predict()`, to find one (1) real-valued feature that yields the lowest mean absolute error (MAE) in the price prediction of an automobile.

1. Append your source code for Part (2) to *AutomobilePricePred_Lab2.py*.
2. Output the lowest MAE real-valued feature as well as the associated MAE.

Deliverable:

All work submitted is subject to the standards of conduct as specified in BCIT Policy 5104. No late assignments will be accepted.

[Sep 16, 2022 @1730] Ensure that your source code is adequately commented and submit *AutomobilePricePred_Lab2.py* to BCIT Learning Hub (Laboratory Submission | Lab 2).