# COMP 4983: Lab Exercise #3

**Mark: /50**

[Due: Sep 23, 2022 @1730
Assignment Submission
Folders]

**Instructions**:

In this lab, you will

- compute the least squares estimate for polynomial regression models on paper for a trivial dataset
- extend Lab Exercise #2 and use Python to
  - ❖ implement a linear regression model to replace the **LinearRegression.fit()** and **LinearRegression.predict()** functions
  - ❖ determine the best combination of two features that will yield the lowest mean absolute error (MAE) in the price prediction of an automobile

## Part 1: Polynomial Regression (on paper)

In this part of the lab, you will compute the least squares estimates for polynomial regression models for a trivial dataset.

Consider a training set consisting of the following three (3) training samples:

| Sample | Value |
|---|---|
| $(x_1, y_1)$ | (2, 4) |
| $(x_2, y_2)$ | (3, 2) |
| $(x_3, y_3)$ | (5, 3) |

1) Determine the value of $\hat{\beta}_0$ and compute the residual sum of squares (RSS) for a zeroth-degree polynomial regression model, i.e., $\hat{y}_i = \hat{\beta}_0$.
2) Determine the values of $\hat{\beta}_0$ and $\hat{\beta}_1$ and compute the RSS for a first-degree polynomial regression model, i.e., $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$.

## Part 2: Lab Exercise #2 Solution

Before you begin this lab, you need to verify the correctness of your implementation from Lab Exercise #2. Use the following metric values to verify your solution:

| | |
|---|---|
| **Mean Absolute Error** | **= 1646.22726** |
| **Root Mean Squared Error** | **= 2253.34207** |
| **Coefficient of Determination** | **= 0.88851** |

In the event that your solution is incorrect, you will need to debug and correct your implementation before proceeding with this lab exercise.

## Part 3: Linear Regression Model Implementation

[30 marks] In this part of the lab, you will replace the **LinearRegression.fit()** and **LinearRegression.predict()** functions used in Lab Exercise #2 by implementing the following functions of a linear regression model, as defined, in your script. Replace ??? with appropriate code:

Function 1:

```
# Fit linear regression model
# Inputs:
#   X: training input
#   y: training output
# Outputs:
#   w: estimated weights of the linear regression model
def comp4983_lin_reg_fit(X, y):
  ???

  return w
```

Function 2:

```
# Predict using the linear regression model
# Inputs:
#   X: test input
#   w: estimated weights from comp4983_lin_reg_fit()
# Outputs:
#   y: predicted output
def comp4983_lin_reg_predict(X, w):
  ???

  return y
```

Steps:

1) Make a copy your submitted *AutomobilePricePred_Lab2.py* from Lab Exercise #2 and rename it as *AutomobilePricePred_Lab3.py* in your working directory.

2) Download the preprocessed automobile price data, *AutomobilePrice_Lab3.csv*, which contains 195 rows and 15 columns, from BCIT Learning Hub (Content | Laboratory Material | Lab 3) and save it in your working directory.

3) Update your script to read from *AutomobilePrice_Lab3.csv* and remove all code previously used for data preprocessing (start and end shown below):

```
# data preprocessing
# remove the 'normalized-losses' column

...
...

# data_subset now should contain 193 rows and 32 columns
print('\n\ndata_subset.info():')
print(data_subset.info())
```

4) Implement the new functions **comp4983_lin_reg_fit(X, y)** and **comp4983_lin_reg_predict(X, w)**. Ensure that these functions are defined before being called.

5) Update your script to call these functions instead of **LinearRegression.fit()** and **LinearRegression.predict()**. You can verify the results of the new functions with that of **LinearRegression.fit()** and **LinearRegression.predict()**.

## Part 4: Feature Selection

[20 marks] In this part of the lab, you will implement a mechanism in your Python script, continuing from Part (3), to determine the best combination of two features from the given preprocessed automobile price data, *AutomobilePrice_Lab3.csv*, that will yield the lowest mean absolute error (MAE) in the price prediction of an automobile. Output the two features and 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 23, 2022 @1730] Ensure that your source code is adequately commented and submit using the filename *AutomobilePricePred_Lab3.py* to BCIT Learning Hub (Laboratory Submission | Lab 3).