Report Assignment 1

1. Write down your regression equation in basic part (1%)
   I. The equation used related to gradient descent:
      guess = m * trainx + b
      error = trainy - guess
      m = m + (error1 * trainx) * learning_rate
      b = b + (error1) * learning_rate

      Where,
      > guess = prediction of the result
      > m is the parameter of w
      > b is the measurement of noise
      > error is the difference between train and prediction result

   II. flag = abs(y1_test[i]-result_y1_test[i])/y1_test[i]
       This equation is to compute the loss

2. Briefly describe the variables you used in the advanced part (1%)
   None

3. Briefly describe the difficulty you encountered (1%)
   a) First time dealing with regression, an extremely time-consuming condition for me, so I only completed the basic part.
   b) The outliers problem causing the results came out at the first time is not that accurate when comparing to the data.
   c) The MAPE of three results quite high.

4. Summarize how you solve the difficulty and your reflections (2%)
   a) Try to understand the linear regression from the lecture slides, and various resources. Then start to implement them to the assignment.
   b) Remove the outliers with the mean and standard deviation of the testing data of every cities as the lower and upper boundaries. In this case, the weird data will be removed from the data set. Then, start to train the data and do predictions of the results.
   c) Change the learning rate for the three data. Each city will have its own learning rate, so that the value of MAPE can be reduced. The MAPE of three data drop below 25% after testing with different learning rates.