National Taiwan University
Institute of Industrial Engineering

Data Analytics

107-2 Homework #03

Due at 09h10, April 3, 2019

Upload one R script file in CEIBA, or send TA and me an email with the attachment.

1. Simulate a "multiple" regression problem by yourself, with sufficient sample size.

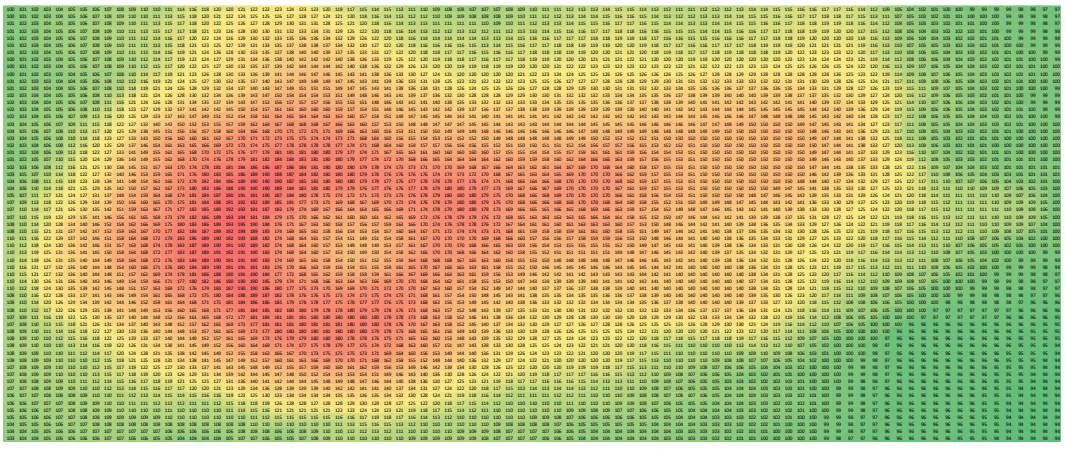
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon, \epsilon \sim N(0, \sigma^2).$$

Use the regression package you can find in your preferable environment (R or Python) to analyze the problem and review the results.

From the perspective of "Machine Learning", let's code the gradient descent method to optimize (minimize) the error function and get the coefficients. Do you get the same results? Demonstrate the evolution of the iterative errors and the searching path in the domain of the error function.

*make any necessary assumptions by yourself if not mentioned above.

2. The volcano dataset can be visualized as the contour below.



We can simply assign the grid coordinates as: $x_1 = \{1, 2, \dots, 87\}; x_2 = \{1, 2, \dots, 61\}$.

Design a mountain climbing algorithm based on *multiple regression models* to arrive the highest point of this volcano, given the starting point at $(87, 1)$, i.e., the right-bottom corner.