Introduction to Data Science
Homework #3

Due: December 4, 2019

This assignment is to practice how to generate a linear regression from a provided numeric dataset, `advertisement.csv`. The data in the dataset are regarding advertisements. We consider the linear regression of a single variable, denoted as $y = w_0 + w_1 x$, among the attributes. There are 5 attributes in the dataset:

1. **Cost**: the cost of the ad
2. **Click**: the number of times the user clicks on the ad
3. **Watch_seconds**: the average number of seconds the user viewed the ad
4. **Ad_seconds**: the total number of seconds of the ad
5. **Dislike**: the number of users who do not like the ad.

Please make regression analysis of the attributes in pairs. E.g., Cost = x is an independent variable and Click = y is a dependent variable. There are ten possible combinations in pairs: (Cost, Click), (Cost, Watch_seconds), (Cost, Ad_seconds), (Cost, Dislike), (Click, Watch_seconds), (Click, Ad_seconds), (Click, Dislike), (Watch_seconds, Ad_seconds), (Watch_seconds, Dislike), and (Ad_seconds, Dislike).

What you have to do are:

1. divide this data set into test or training sets (maybe 40% for testing, 60% for training or other ratios).
2. perform regression analysis to get the coefficient of w_1, the intercept of w_0, and the R-square.
3. plot the regression analysis and each combination is required.
4. allow entering x (say, Cost) to predict y (say, Click) or others.

The related file for this assignment will be put on the *ischool* (`http://www.ischool.ntut.edu.tw/`) platform of school for you to access.

About submitting this homework

- Please upload your homework project named as `HW3-LR-SID.ipynb` to *ischool* platform (`http://www.ischool.ntut.edu.tw/`).
- The **deadline** is the **midnight of December 4**, 2019 and **Late work** is not acceptable.
- Honest Policy: We encourage students to discuss their work with the peer. However, each student should write the program or the problem solutions on her/his own. Those who copy others work will get 0 on the homework grade.