Principles and Applications of Data Science
Homework #3

Due: June 24, 2020

This assignment is to practice how to generate a linear regression from a numeric dataset.We provide the weather data file (`Weather_2009_2016.csv`) for practice. In the dataset, there are twelve attributes; however, we only interest the eleven attributes for exploring the linear regression. These attributes are **Temperature (degC)**, **Pressure (mbar)**, **Tdew (degC)**, **rh (%)**, **VPdef (mbar)**, **sh (g/kg)**, **H2OC (mmol/mol)**, **rho (g/m^3)**, **wv (m/s)**, **max. wv (m/s)**, **wd (deg)**.

The last ten columns are independent variables and the first one is the dependent variable. Assume the linear regression can be denoted as

$$\mathbf{y} = c + a_1\mathbf{x}_1 + \ldots + a_{10}\mathbf{x}_{10}$$

where y is the dependent variable, x_i's are independent variables, c is the constant, and a_i are the coefficients of the linear regression. Please show the coefficients of the linear regression in order (i.e., $c, a_1, a_2, ..., a_{10}$) with the following approaches:

1. Calculate the linear regression from the raw data directly. (You can choose one of the approaches in class for implementation; of course, you must make sure that you won't get a singular matrix if you use the matrix approach.)
2. Generate a *heatmap* for the diagonal correlation matrix with attributes and show your observation.
3. Explore multiple variables with *scatter plot.* The scatter plot of Pandas is a grid of plots of multiple variables one against the other, showing the relationship of each variable to the others. Please state what you observe.
4. Improve the linear regression from question 1 and get a new linear regression if the coefficients are meaningless.

About submitting this homework

- Please upload your homework project named as `HW3-SID.ipynb` to **i-school(Plus)** (`https://istudy.ntut.edu.tw/learn/index.php`) platform .
- The **deadline** is the **midnight of June 24**, 2020 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.