Practice Assignment 1.1: Regression

This Assignment deals with the following topics:

- Linear Regression

The Assignment

In this assignment, you will analyze data in the Air Quality dataset to build a machine learning model that predicts air quality. You will analyze data in the AirQualityUCI.csv file and use functions from the pandas module for loading, inspecting, and querying the data. You are expected to preprocess the data and use the data to train machine learning models from the 'sklearn' library to solve problems.

To begin this assignment, we have provided you with an .ipynb file in Codio, which contains the instructions as inline comments. Moreover, details about the dataset are described in the following sections.

For each question, there are clear instructions in each cell. Follow those instructions and write the code after each block of:

```
# YOUR CODE HERE
raise NotImplementedError()
```

Make sure to delete the line raising an error! Please use the exact variable name if it is specified in the comment.

We'll run a Python test script against your program to test whether the code in each cell is correct.

Dataset Overview:

We will use the Air Quality dataset for this practice assignment. The link to this dataset is shown here:

https://archive.ics.uci.edu/dataset/360/air+quality

The dataset contains 9358 instances of hourly averaged responses from an array of 5 metal oxide chemical sensors embedded in an Air Quality Chemical Multi-Sensor Device. For example, CO(GT) is the hourly averaged concentration CO in mg/m^3, and PT08.S1(CO) represents hourly averaged sensor response (nominally CO targeted), etc.

Date	Time	CO(GT)	PT08.S1(CO)	NMHC(GT)	C6H6(GT)	PT08.S2(NMHC)	NOx(GT)
3/10/2004	18:00:00	2.6	1360	150	11.9	1046	166
3/10/2004	19:00:00	2	1292	112	9.4	955	103
3/10/2004	20:00:00	2.2	1402	88	9.0	939	131
3/10/2004	21:00:00	2.2	1376	80	9.2	948	172

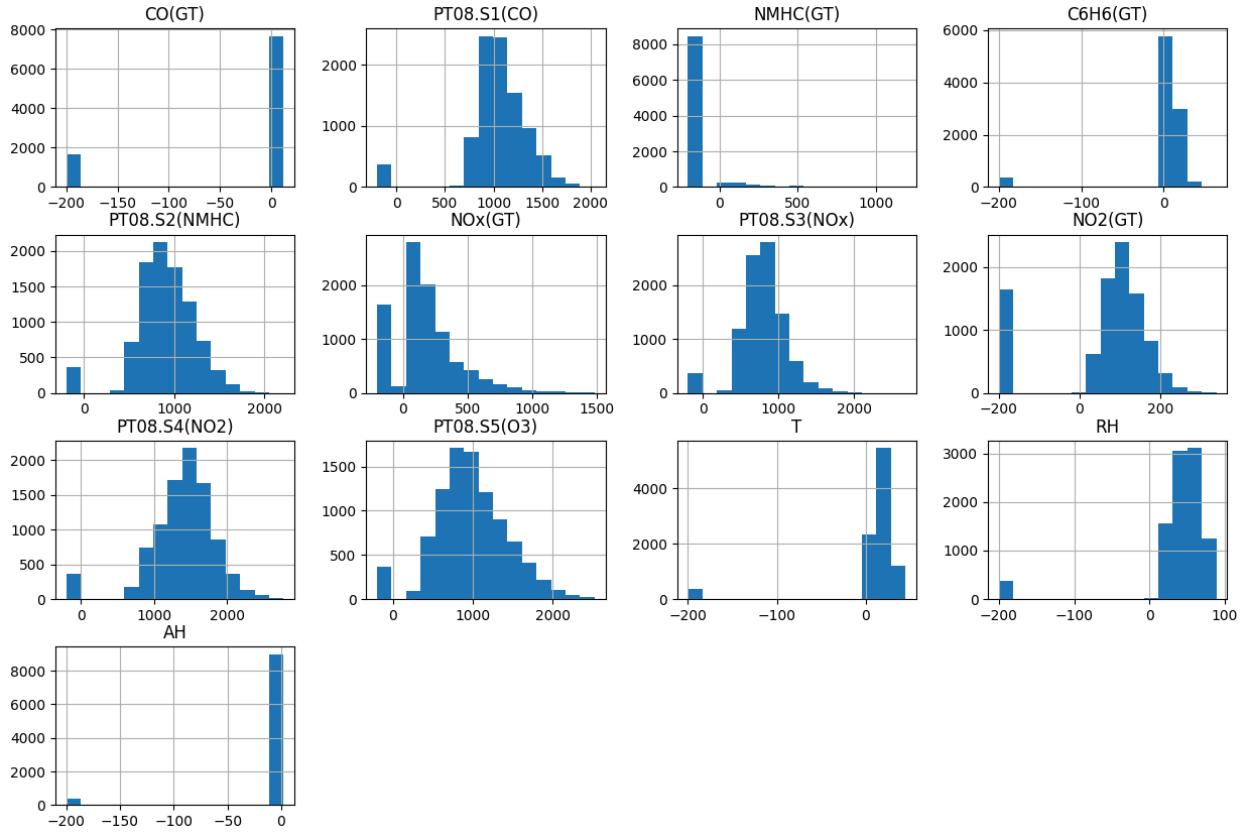
What to submit

Before submitting, click “Validate” to check if your answers are right. When you are satisfied with your answers, in Codio go to the “Education” menu and select “Mark as Completed.”

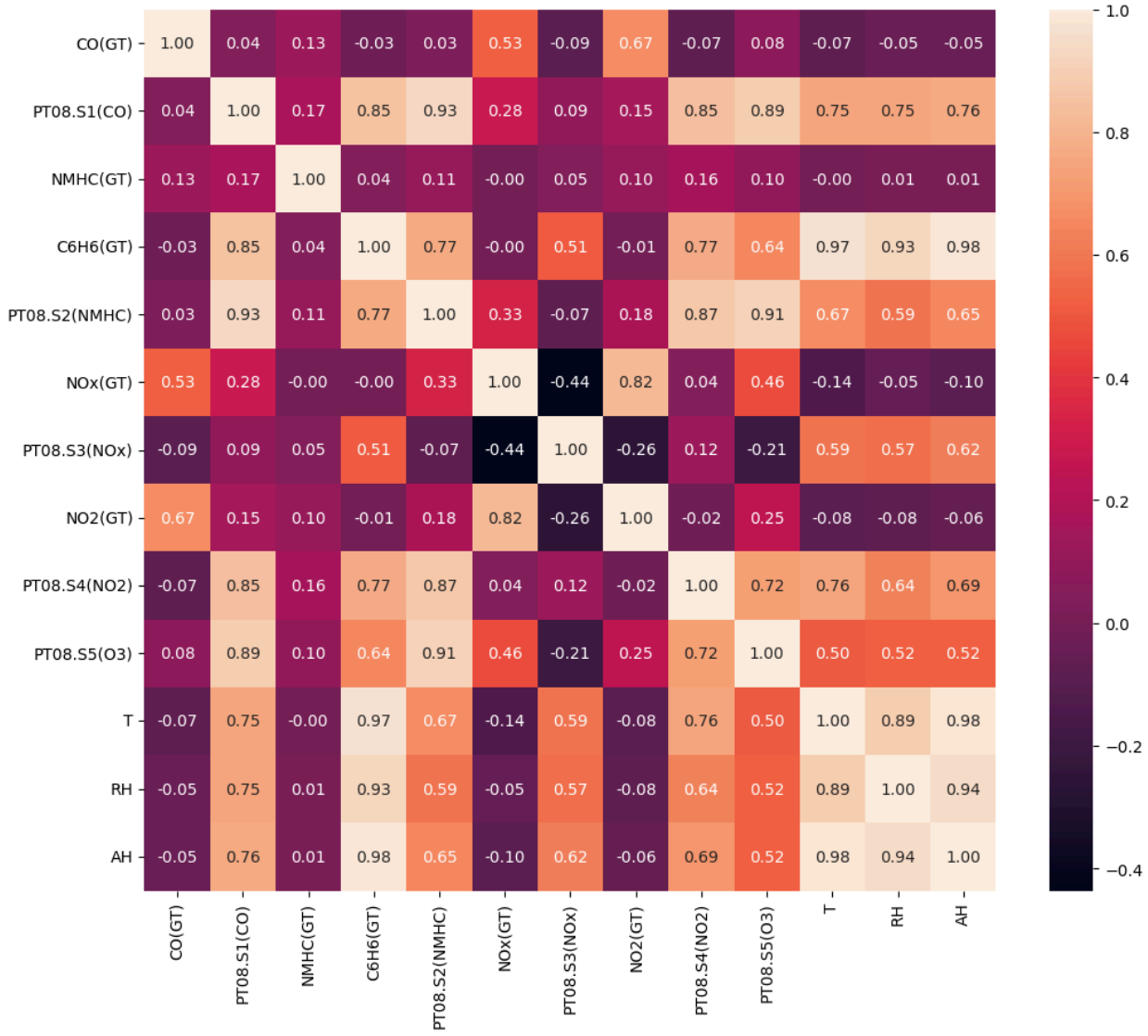
Sample plots:

Here are the sample graphs for Part 03:

1. Histograms:



2. Heatmap:


CO(GT) 1.00 0.04 0.13 -0.03 0.03 0.53 -0.09 0.67 -0.07 0.08 -0.07 -0.05 -0.05
PT08.S1(CO) 0.04 1.00 0.17 0.85 0.93 0.28 0.09 0.15 0.85 0.89 0.75 0.75 0.76
NMHC(GT) 0.13 0.17 1.00 0.04 0.11 -0.00 0.05 0.10 0.16 0.10 -0.00 0.01 0.01
C6H6(GT) -0.03 0.85 0.04 1.00 0.77 -0.00 0.51 -0.01 0.77 0.64 0.97 0.93 0.98
PT08.S2(NMHC) 0.03 0.93 0.11 0.77 1.00 0.33 -0.07 0.18 0.87 0.91 0.67 0.59 0.65
NOx(GT) 0.53 0.28 -0.00 -0.00 0.33 1.00 -0.44 0.82 0.04 0.46 -0.14 -0.05 -0.10
PT08.S3(NOx) -0.09 0.09 0.05 0.51 -0.07 -0.44 1.00 -0.26 0.12 -0.21 0.59 0.57 0.62
NO2(GT) 0.67 0.15 0.10 -0.01 0.18 0.82 -0.26 1.00 -0.02 0.25 -0.08 -0.08 -0.06
PT08.S4(NO2) -0.07 0.85 0.16 0.77 0.87 0.04 0.12 -0.02 1.00 0.72 0.76 0.64 0.69
PT08.S5(O3) 0.08 0.89 0.10 0.64 0.91 0.46 -0.21 0.25 0.72 1.00 0.50 0.52 0.52
T -0.07 0.75 -0.00 0.97 0.67 -0.14 0.59 -0.08 0.76 0.50 1.00 0.89 0.98
RH -0.05 0.75 0.01 0.93 0.59 -0.05 0.57 -0.08 0.64 0.52 0.89 1.00 0.94
AH -0.05 0.76 0.01 0.98 0.65 -0.10 0.62 -0.06 0.69 0.52 0.98 0.94 1.00
CO(GT)
PT08.S1(CO)
NMHC(GT)
C6H6(GT)
PT08.S2(NMHC)
NOx(GT)
PT08.S3(NOx)
NO2(GT)
PT08.S4(NO2)
PT08.S5(O3)
T
RH
AH
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4