Graded Assignment - 11 (PART - B)

The due date for submitting this assignment has passed.
Due on 2024-04-17, 23:59 IST.
You may submit any number of times before the due date. The final submission will be considered for grading.

You have last submitted on: 2024-04-17, 22:06 IST

About the dataset:

dataset URl: https://drive.google.com/file/d/1cBPurUshGKrK72SWsCatRHomjMnAWK6f/view?usp=sharing

The dataset contains 9471 instances of hourly averaged responses from an array of 5 metal oxide chemical sensors embedded in an Air Quality Chemical Multisensor Device. The device was located on the field in a significantly polluted area, at road level,within an Italian city. Data were recorded from March 2004 to February 2005 (one year)representing the longest freely available recordings of on field deployed air quality chemical sensor devices responses. Ground Truth hourly averaged concentrations for CO, Non Metanic Hydrocarbons, Benzene, Total Nitrogen Oxides (NOx) and Nitrogen Dioxide (NO2) and were provided by a co-located reference certified analyzer.

Attributes:

Date (DD/MM/YYYY)
Time (HH.MM.SS)
True hourly averaged concentration CO in mg/m^3 (reference analyzer)
PT08.S1 (tin oxide) hourly averaged sensor response (nominally CO targeted)
True hourly averaged overall Non Metanic HydroCarbons concentration in microg/m^3 (reference analyzer)
True hourly averaged Benzene concentration in microg/m^3 (reference analyzer)
PT08.S2 (titania) hourly averaged sensor response (nominally NMHC targeted)
True hourly averaged NOx concentration in ppb (reference analyzer)
PT08.S3 (tungsten oxide) hourly averaged sensor response (nominally NOx targeted)
True hourly averaged NO2 concentration in microg/m^3 (reference analyzer)
PT08.S4 (tungsten oxide) hourly averaged sensor response (nominally NO2 targeted)
PT08.S5 (indium oxide) hourly averaged sensor response (nominally O3 targeted)
Temperature in Â°C
Relative Humidity (%)
AH Absolute Humidity

Information for questions 1 and 2 and 3.

Load the AirQualityUCI dataset.
Drop 'Date', 'Time', 'Unnamed: 15', 'Unnnamed: 16' columns.
Drop all the rows which has at-least one missing value.

1) What is the shape of the dataset after dropping the rows and colums as mentioned above? ***1 point***

- (9357, 13)
- (9357, 17)
- (9471, 13)
- (9471, 16)

Yes, the answer is correct.
Score: 1

Accepted Answers:

(9357, 13)

2) The task is to predict the absolute humidity at a particular time based on other features. So, we will have AH as the target variable. Under which category does this task fall? ***1 point***

- Regression
- Classification
- Clustering
- None of the above.

Yes, the answer is correct.
Score: 1

Accepted Answers:

Regression

3) What is the value of correlation co-efficient between temperature (T) and relative humidity (RH)?

0.885

Yes, the answer is correct.
Score: 1

Accepted Answers:

(Type: Range) 0.87 , 0.91

1 point

Instructions for questions 4

Separate features and target variable.
split the dataset into training and test dataset in 80:20 proportion with "random_state=1"
Create a pipeline with scaler as StandardScaler and classifier as MLPRegressor.
Classifier should have the following properties:

Classifier should have three hidden layers with 50 neurons each.
set

```
tol=1e-2, alpha=1e-4,solver="adam", learning_rate_init=0.1, max_iter=50, random_state=1
```

4) How many samples are there in the test dataset?

2340

No, the answer is incorrect.
Score: 0

Accepted Answers:

(Type: Numeric) 1872

1 point

5) What is the value of coefficient of determination of the prediction on the training dataset?

0.998

Yes, the answer is correct.
Score: 1

Accepted Answers:

(Type: Range) 0.96 , 1.03

1 point

6) What is the value of coefficient of determination of the prediction on the test dataset?

0.999

Yes, the answer is correct.
Score: 1

Accepted Answers:

(Type: Range) 0.96 , 1.03

1 point

7) What will be the prediction of the model on the first intance of training set?

0.7626

No, the answer is incorrect.
Score: 0

Accepted Answers:

(Type: Range) 1.57 , 1.61

1 point