Graded
Assignment

● Practice Assignment 8.2 - Not
Graded
Assignment

● Graded Assignment - 8 (PART - A)
Assignment

● Graded Assignment - 8 (PART - B)
Assignment

○ Week 9 ⌄

○ Week 10 ⌄

○ Week 11 ⌄

○ Week 12 ⌄

○ MOCK Unit ⌄

○ Supplementary Contents ⌄

○ Answer Key for Exam ⌄

Graded Assignment - 8 (PART - A)

The due date for submitting this assignment has passed.
Due on 2024-03-17, 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: 2024-03-17, 22:07 IST

(Common data for Q1 to Q5 graded questions)

Step 1: Download the dataset using following link: (https://drive.google.com/file/d/1v-uxWEgTl0GDCOTZOX3shUMkTf1a_CL7/view?usp=sharing)

Step 2: Import the data in google colab using pd.read_csv().

Step 3: Seperate features and target data in seperate variable X and Y.

Step 4: Convert dataframe X and series y into array and save it in variable X_array,y_array.

Step 5: Split the dataset using train_test_split. (Keep parameter test_size=0.3 and random_state=10).

Step 6: Reshape the dataset in such a way that each entry of data has 90 samples.

Step 7: Use SGD regressor as an estimator and partial_fit to fit the dataset on the model. (Set random_state=10)

Step 8: Calculate different evaluation metrics value like mean_square_error, R2_score.

Use the training set for fitting the model and use the test data to make the predictions.

Note: No need to scale the data. It's already scaled.

Answer the below questions.

---

1) How many fetaures are there in the dataset? **1 point**

○ 10
○ 11
○ 90,000
○ 9000

Yes, the answer is correct.

Score: 1

Accepted Answers:

10

---

2) What is the value of intercept you got after training the model using SGDRegressor?(select the closest answer) **2 points**

○ +0.005
◉ -0.005
○ +0.105
○ -0.105
○ 0

Yes, the answer is correct.

Score: 2

Accepted Answers:

-0.005

---

3) What is the value of cofficient corresponding to "feature-3"you got after trainig the model using SGDRegressor? (select the closest answer)

81.237

Yes, the answer is correct.

Score: 2

Accepted Answers:

(Type: Range) 81 , 82

2 points

---

4) What is the value of R2 score for test data

0.999

Yes, the answer is correct.

Score: 3

Accepted Answers:

(Type: Range) 0.99 , 1

3 points

5) What is the value of cofficient corresponding to "feature-5" after 5th iteration.

52.09

No, the answer is incorrect.

Score: 0

Accepted Answers:

(Type: Range) 55,65

3 points

(Common data for Q6 to Q8 Graded questions)

This dataset was constructed by adding elevation information to a 2D road network in North Jutland, Denmark (covering a region of 185 x 135 km^2). Elevation values where extracted from a publicly available massive Laser Scan Point Cloud for Denmark. This 3D road network was eventually used for benchmarking various fuel and CO2 estimation algorithms. This dataset can be used by any applications that require to know very accurate elevation information of a road network to perform more accurate routing for eco-routing, cyclist routes etc. For the data mining and machine learning community, this dataset can be used as 'ground-truth' validation in spatial mining techniques and satellite image processing. It has no class labels,

Use this dataset to guess some missing elevation information for some points on the road.

Column names:

OSM_ID: OpenStreetMap ID for each road segment or edge in the graph.

LONGITUDE: Web Mercaptor (Google format) longitude

LATITUDE: Web Mercaptor (Google format) latitude

ALTITUDE: Height in meters.

Load the dataset from link("https://archive.ics.uci.edu/ml/machine-learning-databases/00246/3D_spatial_network.txt"). Set parameter chunk size=20000 and iterator =True in pd.read_csv().

NOTE: The above file doesn't have column names

Scale your whole dataset first with standard scalar using partial_fit method. Then use SGDRegressor(random state=10) on the dataset and answer the following.

6) Check how many total samples are there in the dataset?   *1 point*

○ 470000

○ 400000

○ 90,000

◉ 434874

Yes, the answer is correct.

Score: 1

Accepted Answers:

434874

7) What is the value of intercept after 7th iteration. (select the closest option).   *2 points*

◉ 21.3

○ 15.0

○ 10.8

○ 5

Yes, the answer is correct.

Score: 2

Accepted Answers:

21.3

8) What is the value of the coefficient corresponding to the longitude feature after the 7th iteration? (select the closest option).   *2 points*

○ [9.6]

◉ [4.7]

○ [8.6]

○ [5.5]

○ [1.7]

Yes, the answer is correct.

Score: 2

Accepted Answers:

[4.7]

(Common data for Q9 to 11)

Load Iris datset on Colab and use KNN classifier to build the model Using following steps.

Step 1: Load the dataset and split it using train_test_split by keeping: test_size= 0.2 random_state=10

Step 2: Use Normalizer() as a scaling function to scale the data.

Step 3: Use KNeighborsClassifier(K) as an estimator to predict the output.

9) Which of the following K value gives the best accuracy on test set.   *1 point*

○ k=4

○ k=2

○ k=3

◉ All K value given in the option gives same score.

Yes, the answer is correct.
Score: 1

Accepted Answers:

All K value given in the option gives same score.

10) What is the accuracy for k=3?   *1 point*

◉ 0.96

○ 0.91

○ 0.99

○ 1.0

Yes, the answer is correct.
Score: 1

Accepted Answers:

0.96

11) Compute wieghted F1 score value for k=3.(Keep parameter average='weighted')

0.967

Yes, the answer is correct.
Score: 1

Accepted Answers:

(Type: Range) 0.95 , 0.98

1 point