☰   DSCI-6601-001 (Pract Machine Learning 77223)   ✉ 💬 🔔   SA

Assignment 2

Instructions

Total Marks: [10]

For this assignment, we will work primarily with the EPA fuel economy dataset, which is provided for download in this sub-module in .CSV format.

For the questions below, create either a standalone Python program (.py) or an iPython Notebook (.ipynb) to implement code based on the directions to follow:
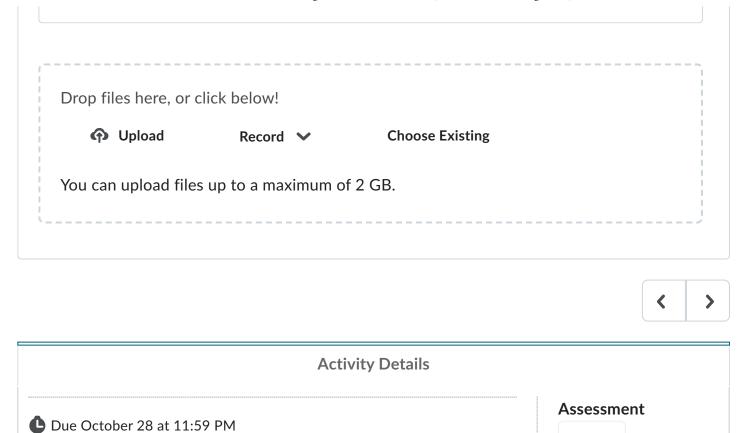
1. (1 Marks) Load the given EPA fuel economy data set into a pandas dataframe. Write code which removes the non-numeric columns from the data frame. Create a new column named "CityHighwayDiff" which is calculated by taking the difference of the "city08" column and "highway08" column. Plot a histogram of this column.

2. (3 Marks). Using the EPA fuel economy data set. Consider the "year" column. Create a new column "Decade" from the year column which maps each year into a class based on its decade. Plot a histogram of the Decade column to visualize the occurrences in each of the decades. Remove any other non-numeric columns from the dataframe. With the columns remaining, create a linear classifier that predicts the decade of a particular car based on the rest of the columns. Train the classifier using the entire dataset. Calculate and print the training error of the resulting classifier.

3. (3 Marks) Building on 3, split the training set used into a training and test set (split evenly 50:50). Re-fit the linear classifier developed above on the 50% training set and calculate and print the mean absolute error of the predictor using the 50% test set.

4. (3 Marks) Pick 2 of the numeric columns of the dataframe to eliminate from the table. Re-train the classifier using this new dataset. Compare the testing and training error using the 50/50 split of the missing columns classifier with that containing the full set of columns.

Submissions

Assignment 2 6601 Final.ipynb (92.39 KB)   Oct 28, 2022 9:06 PM

Drop files here, or click below!

⤒ Upload   Record ⌄   Choose Existing

You can upload files up to a maximum of 2 GB.

‹   ›

Activity Details

🕐 Due October 28 at 11:59 PM

Assessment

10 / 10

Last Visited Oct 26, 2022 4:04 PM