Cars Data - (Assignment 4 - Question 4)

- subject Machine Learning / AI

DESCRIPTION

Consider an automobile data set with values such as Name of the car Model, its mileage, number of cylinders, number of gears and so on.

Here's a preview of the data under consideration:

model	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21	6	160	110	3.9	2.62	16.46	0	1	4	4
Mazda RX4 Wag	21	6	160	110	3.9	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.32	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.44	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.46	20.22	1	0	3	1
Duster 360	14.3	8	360	245	3.21	3.57	15.84	0	0	3	4

The data is present in the file named **mtcars.csv** which is present at the location **/data/training/mtcars.csv**

Write a Python code to calculate the difference between the means 10-fold cross validation scores of ridge regression and lasso regression with alpha as 1.0

Hint: Perform steps as mentioned below:

- Load data
- Use all the columns except 'mpg' & 'model' as predictors (x)
- Use 'mpg' as the response column(y)
- Perform Lasso & Ridge regression on this data. Use all the rows as training data for performing regression.
- Perform cross validation on both the regression results for the above x & y with cv=10 and default scoring
- Calculate and ***print*** the difference by subtracting the **mean score of ridge** from the **mean score of lasso** regression

Input Format:

Read the input file **/data/training/mtcars.csv**

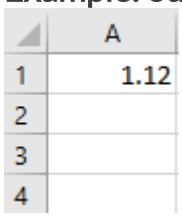
Output Format:

- You have to perform the operations as described above and write the value of difference in the mean scores as stated above in a file named **output.csv,** which should be present at the location **/code/output/output.csv**

- **output.csv** should contain the value rounded to 2 decimals in the first row

  Sample Output:

  Example: output.csv will have data looking like this:

	A
1	1.12
2	
3	
4	

  DATASETS

- Training dataset