CS422/622- HW 3

In HW3, we are going to write codes for a linear regression model from scratch to solve a regression problem: to predict mpg.

Download data from https://archive.ics.uci.edu/ml/datasets/auto+mpg, where there are 9 attributes below:

1. mpg: continuous
2. cylinders: multi-valued discrete
3. displacement: continuous
4. horsepower: continuous
5. weight: continuous
6. acceleration: continuous
7. model year: multi-valued discrete
8. origin: multi-valued discrete
9. car name: string (unique for each instance)

You can remove the last column of "car name", which is a unique value for each instance. The regression problem is to predict "mpg" values using the other variables. So, the regression will include 7 independent variables (2—8) and one dependent variable (1). You must use 10-fold cross-validation. For each fold, you must provide the table below:

1. Coefficients of seven independent variables.
2. Root Mean Square Error (RMSE): The equation of RMSE is $\sqrt{\sum_{i=1}^{N}(y_i - \mathbf{X_i b})^2}$, where y_i is a ground truth value of the i-th test data, $\mathbf{X_i b}$ is a prediction from the linear regression model, and *N* is a number of test data.

	cylinders	displacement	horsepower	weight	acceleration	model year	origin	RMSE
Fold 1								
Fold 2								
Fold 3								
…	…	…	…	…	…	…	…	…
Fold 10								

Submission:

You must submit the followings to WebCampus:

1. MS word file
   - Describe what you did for the homework assignment.
     - Must include the table of the 10-fold CV

2. Source code file(s)
    - Must be well organized (comments, indentation, …)

You must submit the files SEPERATELY. DO NOT compress into a ZIP file. If you fail to provide all required information or files, you may be given zero score without grading.

Rubric:

- If used a library for linear regression instead of writing code from scratch, zero will be given.
- You can use any functions or libraries other than library of linear regression. E.g., libraries for cross-validation are okay.
- No result table on the word file will deduct at least 50 points.

Deadline:

You must submit HW3 by **Friday, March 24, 2023**. Late assignments will not be accepted.