

C S 487/519 Applied Machine Learning

Compare regression methods

1 Objective

In this *individual* homework, you are required to understand and compare several regression algorithms.

2 Requirements

2.1 Tasks

- (1) (60 points) Write code to conduct regression by
 - (a) (35 points) utilizing several regression functions: (i) LinearRegression, (ii) RANSACRegressor, (iii) Ridge, and (iv) Lasso. These functions are provided by the Python scikit-learn library.
 - (b) (25 points) using one approach to conduct non-linear regression, and
- (2) (20 points) Each regressor needs to be tested using the California housing dataset, which can be loaded using `fetch_california_housing` from `sklearn.datasets`. You need to use all the columns and all the instances in this dataset. During lectures, I may just use one column and a subset of the dataset for demonstration purpose. If you decide to use fewer columns/instances of the dataset, you need to show an analysis of why you chose to use only a subset of features/instances. For example, you can provide a correlation analysis to justify why you are not using all columns, or show a random sample of the instances. If you do not justify why you are using fewer columns, points will be deducted.
- (3) (15 points) Properly analyze the regressors' behavior by applying the knowledge that we discussed in class. Such analysis should include at least Mean squared error (MSE) (or R2 score, or residual plots) and running time. Put the analysis to `report.pdf` file.
- (4) (5 points) Write a readme file `readme.txt` with detailed instructions to run your program.

2.2 Other requirements

- Your Python code should be written for **Python version 3.5.2 or higher**.
- Please write proper **comments** in your code to help the instructor and teaching assistants to understand it.
- Please properly organize your Python code (e.g., create proper classes, modules).
- You can put your code to Jupyter Notebook or a `.py` file.

3 Submission instructions

Put all your files (Python code, readme file, report, etc.) to a zip file named `hw.zip` and upload it to Canvas.

4 Grading criteria

- (1) **ZERO point will be given if your code does not work. Please do not submit code that you did not test and make sure it works.**
- (2) The score allocation has been put beside the questions.
- (3) FIVE points will be deducted if files are not submitted in the required format.
- (4) If the total points are more than 100. Your grades will be scaled to the range of [0,100].

- (5) Please make sure that you test your code thoroughly by considering all possible test cases. For this homework, your code will NOT be tested using more datasets. Thus, it does not need to be flexible to accept different datasets as input.