

CS 489/689
INTRODUCTION TO MACHINE LEARNING
FALL 2020
Assignment #2

Due Date/Time: 10/09/2020 @ 11:59PM
Total Points: 100

CS 489 students may complete this assignment individually or in teams of two; CS 689 students are to complete it individually.

Description:

- For this assignment, you will implement two *linear regression algorithms* in Python, Julia, or MATLAB to solve a *regression problem*. You may use any and all built-in functions including linear regression functions.
- **Ordinary least squares (OLS) solution:** Implement the ordinary least squares solution as discussed in class.
- **Linear regression with gradient descent:** Implement gradient descent for linear regression as discussed in class.
- For both algorithms, don't forget to include the bias term in the parameter vector.
- Train and test your model with a dataset of your choosing that meets the following criteria:
 - Number of features: 3+
 - Feature characteristics: Real-valued
 - Output characteristics: Real-valued
- Use 80% of the dataset for training and 20% for testing your model.
- If you'd like, you may use one of the following sources to find a dataset:
 - University of California, Irvine Machine Learning Repository <https://archive.ics.uci.edu/ml/index.php>
 - Kaggle <https://www.kaggle.com/>
 - Awesome Public Datasets <https://github.com/awesomedata/awesome-public-datasets>
 - Google Dataset Search Engine <https://datasetsearch.research.google.com/>
 - Microsoft Research Open Data <https://msropendata.com/>
 - U.S. Government's Open Data <https://www.data.gov/>
 - Registry of Research Data Repositories <https://www.re3data.org/>
 - CMU Libraries <https://guides.library.cmu.edu/machine-learning/datasets>
- Summarize your approach and results in a report that includes at least the following:
 - The dataset you used, its source and characteristics.

- The data preprocessing steps you took (if any).
 - The solution $\hat{\mathbf{w}}$ for both algorithms.
 - The learning rate(s) you used for gradient descent and how many iterations it took for gradient descent to converge.
 - Relevant evaluation metrics for the training dataset for both algorithms.
 - Relevant evaluation metrics for the test dataset with for both algorithms.
 - Any additional details you would like to include.
- Submit your report along with your dataset and source code. Feel free to include your code in the report, but you also need to submit your source code files (.py, .jl, or .m) and your dataset separately, so that your results can be replicated for scoring.

Submission Instructions:

Compress all the files and name the submission file **<YourLastName>_Assignment2:**

- If you are completing the assignment individually, your last name is Smith, and you are submitting a .zip file, the file should be named Smith_Assignment2.zip.
- If you are completing the assignment as a team of two, your last names are Rogers and Smith, and you are submitting a .zip file, the file should be named Rogers_Smith_Assignment2.zip. Only one of the team members needs to submit the assignment.

I will set Canvas to allow unlimited number of submissions and will only grade the last submission. So, please do not wait until the last minute to submit as you can always submit an updated version.