CS4990 Fall 2019 Project Assignment 1

Total points: 100

Due date: Thursday, October 3, 2019

Purpose:

1. Warm up your Python programming skills.
2. Understand linear regression and gradient descent.

Task Description:

In this assignment, you will use linear regression to predict house price in Boston suburbs. An iPython notebook is provided ("Assignment 1 – Part 1.ipynb"), where some of the early steps that prepare the training data and validation data have been implemented for you. (Note: due to the simplicity of this linear regression example, we simply treat testing data as validation data.)

Please use Google's Colab to implement the tasks 1-4 and prepare a report for task 5:

- **Task 1: Use** LinearRegression() in Scikit-learn library.
- **Task 2: Implement** analytical solution (based on closed form of the optimal solution given in slides) to perform linear regression.
- **Task 3: Implement** basic gradient descent to perform linear regression. Please tune the parameters to get close to the accuracy of the linear regression model from scikit-learn library. Also, **Use** matplotlib to plot the learning curves showing how training error and validation errors along iterations.
- **Task 4: Implement** stochastic gradient descent method to perform linear regression. Please tune the parameters to get close to the accuracy of the linear regression model from scikit-learn library. Also, **Use** matplotlib to plot the learning curves showing how training error and validation errors along batches.
- **Task 5: Write** a report to **analyze and discuss** the impact of different **batch sizes** on the performance of the models trained in Task 4.

In each of the tasks 1 - 4, you need to

- Display the resulting weights (intercept and coefficients)
- Display the resulting error $\ell(w) = \frac{1}{2N}\sum_{i=1}^{N}[t^{(i)} - y(x^{(i)})]^2$ on training data and validation data, respectively. (Please note the difference between our loss function and the one in Scikit-learn)

What to Submit?

1. A completed iPython notebook for tasks 1-4 (Note: properly comment each code line)
2. A report for task 5.
3. Please zip them into a file (yourname_assignment1.zip) and submit the zipped file in blackboard