Indian Institute of Technology Bombay
EE 769 - Introduction to Machine Learning
Assignment 1 Statement -- Linear Regression from Scratch -- January 12, 2023

Instructions:

- Prepare an ipynb and a py file, and name it A1_<RollNo>.ipynb/py and submit both by January 26, 2023 on Moodle.
- You can discuss the concepts with the other students, but each student should write their own code and comments.
- Write copious comments. Each line should ideally have a trailing comment to explain what it does. Before each code block a text block should explain the intent of the trailing code block. Include a trailing text block to note your observations.
- Include some kind of unit testing in code blocks that do not have an output, such as those defining functions alone.
- Max points 55, which will be scaled to approximately 8% towards course total.

Tasks:

1. Write a function to generate a data matrix X. Inputs: Number of samples, feature dimension. Output: Data matrix X. [1]
2. Write a function to generated dependent variable column t. [1]
    a) Inputs: Data matrix X, weight vector for each column, bias w0, noise variance
    b) Output: Target vector t
3. Write a function to compute a linear regression estimate. [1]
    a) Input: data matrix X and weight vector w
    b) Output: y
4. Write a function to compute the mean square error of two vectors y and t. [1]
5. Write a function to estimate the weights of linear regression using pseudo-inverse, assuming L2 regularization [2]:
    a) Input: X, t, and lambda
    b) Output: w, MSE, y
6. Write a function to compute the gradient of MSE with respect to its weight vector. [2]
    a) Input: X matrix, t vector, and w vector
    b) Output: gradient vector
7. Write a function to compute L2 norm of a vector w passed as a numpy array. Exclude bias w0. [1]
8. Write a function to compute the gradient of L2 norm with respect to the weight vectors. [2]
    a) Input: X matrix and w vector
    b) Output: gradient vector, where gradient with respect to w0 is 0.
9. Write a function to compute L1 norm of a vector w passed as a numpy array. Exclude bias w0. [1]
10. Write a function to compute the gradient of L1 norm with respect to the weight vectors. [2]
    a) Input: X matrix and w vector
    b) Output: gradient vector, where gradient with respect to w0 is 0.
11. Write a function for a single update of weights of linear regression using gradient descent. [2]
    a) Input: X, t, w, eta, lambda 2, lambda 1. Note that the weight of MSE will be 1
    b) Output: updated weight and updated MSE
12. Write a function to estimate the weights of linear regression using gradient descent. [3]
    a) Inputs: X, t, lambda2 (default 0), lambda1 (default 0), eta, max_iter, min_change_NRMSE
    b) Output: Final w, final RMSE normalized with respect to variance of t.
    c) Stopping criteria: Either max_iter has been reached, or the normalized RMSE does not change by more than min_change_NRMSE
13. Run multiple experiments (with different random seeds) for, plot the results of (box plots), and comment on the trends and potential reasons for the following relations:
    a) Training and validation NRMSE obtained using pseudo inverse with number of training samples [2]
    b) Training and validation NRMSE obtained using pseudo inverse with number of variables [2]
    c) Training and validation NRMSE obtained using pseudo inverse with noise variance [2]
    d) Training and validation NRMSE obtained using pseudo inverse with w0 [2]
    e) Training and validation NRMSE obtained using pseudo inverse with lambda2 [2]
    f) Time taken to solve pseudo inverse with number of samples and number of variables and its breaking points [2]
    g) Training and validation NRMSE obtained using gradient descent with max_iter [2]
    h) Training and validation NRMSE obtained using gradient descent with eta [2]
    i) Time taken to solve gradient descent with number of samples and number of variables and its breaking points [2]
    j) Time taken to solve gradient descent with number of variables and its breaking point [2]
    k) Training and validation NRMSE and number of nearly zero weights obtained using gradient descent with lambda2 [2]
    l) Training and validation NRMSE and number of nearly zero weights obtained using gradient descent with lambda1 [2]
    m) Training and validation NRMSE for optimal lambda2 with noise variance [2]
    n) Training and validation NRMSE for optimal lambda1 with noise variance [2]
    o) Experiment (f) but, this time with number of training samples and number of variables [2]
14. Write your overall learning points by doing entire assignment. [4 + 2 bonus for diligent work and critical thinking]
15. Quote your references, including roll numbers of fellow students with whom you discussed. Be specific about which part was inspired by what source or which friend.