EE769 Introduction to Machine Learning (Jan 2024 edition)
Electrical Engineering, Indian Institute of Technology Bombay
Assignment – 1 : Supervised Learning from Scratch

Instructions:

a) The assignment is worth 12 points.
b) Submit a single ipython notebook on Moodle before the deadline. You are allowed to consult friends, the internet, and LLM; but copying and pasting code is discouraged. Copying and pasting with trivial modifications and no citation will lead to an FR or NOAUDIT grade with no further chances.
c) Use only basic matrix arithmetic and linear algebra functions from numpy and plot function from matplotlib libraries in python for this assignment.
d) Insert a comment at the end of each line of code to mention its inspirational source number and its purpose. Any line without a comment will be considered copied without citation and will lead to an FR or NOAUDIT grade.
e) Between code cells, insert text cells to note down your observations and to demonstrate your critical thinking abilities, as well as your intention for the next code cell. Your observations will be graded.
f) Use good coding practices such as avoiding hard-coding, using self-explanatory variable names, using functions (if applicable). Coding style will also be graded.
g) Add a text block numbered reference (inspirational sources) as asked in (d) above -- friend's roll no., ChatGPT, internet link 1, etc.

Problem statements:

1. Write a function to generate an input data matrix **X** of size NxD for regression. [0.5]
   a) Input: Sample size N and a generator matrix **S** of size MxD
   b) Working: First generate a random 2-D array of size NxM where each column has a standard normal distribution and is independent of the other columns. Then multiply this with the generator matrix **S** of size MxD to give an output matrix **X** of size NxD. The idea here is that if the generator matrix **S** of size MxD is an identity matrix, then each column of **X** will remain independent; otherwise we can introduce correlations in the matrix columns of **X**.

2. Write a function to generate the target vector **t** of size Nx1: [0.5]
   a) Input: Data matrix **X** of size NxD, weight vector **w** of size D+1 and noise variance σ
   b) Working: Check for dimension mismatch between **X** and **w**, multiply **X** with **w** (sans one element) and add the bias (the excluded element), then add zero-mean Gaussian noise with variance σ.

3. Examine the behavior of the analytical solver based on pseudo-inverse (pinv) in numpy.linalg package with respect to the size of the data matrix. Plot a graph of the time taken with respect to N (use log scale for both axes), with D fixed to 10. Is there any strange behavior in time taken to solve the problem above a particular value of N? What could be the reason for the same? [1]

4. Write a function to calculate the normalized root mean squared error (NRMSE) between a target vector **t** and a predicted vector **y**. [0.5]

5. Write a function to calculate gradient of mean squared error (MSE) with respect to weights of linear regression. Figure out what should be the inputs and outputs. [0.5]

6. Write a function to calculate gradient of L2 norm of weights with respect to weights. [0.5]

7. Write a function to calculate gradient of L1 norm of weights with respect to weights. [0.5]

8. Write a function to perform gradient descent on MSE + λ_1 L1 + λ_2 L2 for linear regression. Use an appropriate stopping criterion. [1]

9. Examine the impact of σ on the NRMSE for linear regression using gradient descent. Average the results of the following experiment run five times for each value of σ where **G** is an identity matrix. Generate a random data matrix **X** and target vector **t** with noise variance σ, and split it into training and validation sub-matrices and sub-vectors. Train using gradient descent on training subset, and test on the validation subset. Plot average NRMSE on validation subset for five runs versus σ. Comment on the results. [1.5]

10. Examine the impact of N and λ_2 on the NRMSE for linear regression using gradient descent. Create lists of N and λ_2 values (use log scale, 5 each, 25 pairs). Average the results of the following experiment run five times for combination of N and λ_2 value pair for a fixed generator matrix **G** and noise variance σ. Comment on the results. [1.5]

11. Examine the impact of λ_1 on variable elimination. Generate a single data matrix **X** and plot weights versus $1/\lambda_1$. Comment on the results. Introduce correlations in the columns of **X** and repeat the experiment. Are the results different? Comment on the results. [1]

12. Show the grouping effect of elastic net on correlated columns of **X**. [1]

13. Write a function for generating linear binary classification vector **t** with noise variance σ. [0.5]

14. Write a function for computing gradient of binary cross-entropy for logistic regression. [0.5]

15. Repeat experiment 10 for binary classification. [1]