CARNEGIE MELLON UNIVERSITY 07-280

HOMEWORK 5

DUE: Thursday, Feb. 19, 2025

https://www.cs.cmu.edu/~07280

INSTRUCTIONS

- **Format:** Use the provided LaTeX template to write your answers in the appropriate locations within the *.tex files and then compile a pdf for submission. We try to mark these areas with STUDENT SOLUTION HERE comments. Make sure that you don't change the size or location of any of the answer boxes and that your answers are within the dedicated regions for each question/part. If you do not follow this format, we may deduct points.

  You may also digitally annotate the pdf. Illegible handwriting will lead to lost points. However, we suggest that try to do at least some of your work directly in LaTeX.

- **How to submit written component:** Submit to Gradescope a pdf with your answers. Again, make sure your answer boxes are aligned with the original pdf template.

- **How to submit programming component:** See section Programming Submission for details on how to submit to the Gradescope autograder.

- **Policy:** See the course website for homework policies, including late policy, and academic integrity policies.

Name	
Andrew ID	
Hours to complete all components (nearest hour)	

1 [12 pts] Programming

1. **[2 pts]** Programming Q1a

   Provide the two plots of the fitted gradient descent models (left box) and the error rate of the models over training (right box).

   Your Answer

   Your Answer

2. **[3 pts]** Programming Q1b

   After performing training, given the images you provided above, in what order did the models converge? Using your knowledge of gradient descent and empirical loss minimization, explain why they converged in that order in one sentence. (Note: the learning rate and the number of epochs are the same for all 3 models during their training).

   Order (first to last)

   Explanation

3. **[4 pts]** Programming Q2a

   Compute the big-O complexity of calculating the closed form solution to weighted linear regression and the complexity of calculating the gradient for **one iteration** of weighted linear regression trained via gradient descent given the following:

   - N by M design matrix X (N samples and M features)
   - a length N vector y of labels
   - an N by N diagonal matrix R

   Assume that we can invert an n by n matrix in $O(n^3)$. Also recall that multiplying by a diagonal matrix is less-complex than normal matrix multiplication. Lastly, ensure that you are considering an optimal order of operations when multiplying matrices/vectors (matrix multiplication is associative).

Closed-form	Gradient descent (1 iteration)

4. **[1 pts]** Programming Q2b

   Given your answer for the time complexities of the two algorithms above, when the number of features, M, is large, which training method is preferred for weighted linear regression?

   ◯ Closed-Form

   ◯ Gradient Descent

5. **[2 pts]** Programming Q3

   Provide the plot of the best model on the energy usage dataset. The code to create the plot has been given to you.

Your Answer

2 Collaboration Policy

After you have completed all other components of this assignment, report your answers to the following collaboration questions.

1. Did you receive any help whatsoever from anyone in solving this assignment? If so, include full details including names of people who helped you and the exact nature of help you received.

Your Answer

2. Did you give any help whatsoever to anyone in solving this assignment? If so, include full details including names of people you helped and the exact nature of help you offered.

Your Answer

3. Did you find or come across code that implements any part of this assignment? If so, include full details including the source of the code and how you used it in the assignment.

Your Answer