Weekly Homework 2

Dr. Sujit Das
Deep Learning

Exercise 1. . Nesterov Accelerated Gradient (NAG)

Given the function:

$$f(x) = x^2 + 5x + 6$$

and a starting point:

$$x_0 = 0$$

find the value of x_1 using the Nesterov Accelerated Gradient (NAG) algorithm.

Assume:

$$\eta = 0.1 \quad \text{(learning rate)}, \quad \gamma = 0.9 \quad \text{(momentum coefficient)}$$

Exercise 2. . Stochastic Gradient Descent (SGD)

Suppose you have a dataset with three data points, and the loss function for each point is given by:

$$L_1(w) = (w-1)^2$$
$$L_2(w) = (w-2)^2$$
$$L_3(w) = (w-3)^2$$

Starting with an initial weight:

$$w_0 = 0$$

What would be the value of w_1 after the first iteration of SGD, assuming a learning rate $\eta = 0.1$?

For this problem, randomly select the first data point $L_1(w)$ to update the weight.

Exercise 3. . Mini-Batch Stochastic Gradient Descent

Using the same loss functions from Problem 2:

$$L_1(w) = (w-1)^2, \quad L_2(w) = (w-2)^2, \quad L_3(w) = (w-3)^2$$

What is the value of w_1 after the first iteration of Mini-Batch SGD?

Given:

$$w_0 = 0, \quad \eta = 0.1, \quad \text{batch size} = 2$$

Assume the mini-batch consists of the first two data points $L_1(w)$ and $L_2(w)$.

Guidelines for the students

Objective

The objective of this assignment is to help students:

- Understand and apply the concept of gradient-based optimization algorithms such as Stochastic Gradient Descent (SGD), Mini-Batch SGD, and Nesterov Accelerated Gradient (NAG).
- Analyze the impact of learning rate, momentum, and batch size on parameter updates.
- Gain hands-on experience in performing one-step updates of optimization algorithms using given loss functions.
- Strengthen mathematical intuition behind loss minimization and convergence behavior of gradient descent methods.

Expected Learning Outcomes

Upon successful completion of this assignment, students will be able to:

- Perform parameter updates using Stochastic Gradient Descent (SGD), Mini-Batch SGD, and Nesterov Accelerated Gradient (NAG) for given loss functions.
- Demonstrate understanding of gradient-based optimization principles through step-by-step calculations.
- Analyze how hyperparameters like learning rate, momentum, and batch size affect optimization.
- Build foundational knowledge for implementing advanced optimization techniques in deep learning frameworks.
- Interpret and explain the mathematical intuition behind optimization steps and convergence.

Instructions

1. Attempt all questions unless specified otherwise.
2. Show all steps clearly:
    - Forward pass calculations.
    - Loss computation.
    - Backpropagation with proper chain rule application.

3. Include all formulas used in each step.
4. Use neat handwriting (if handwritten) or typeset (LaTeX preferred, optional).
5. **No direct ChatGPT or AI-generated answers allowed. Use your own understanding.**

Submission Deadline

[11-08-2025]

Late submissions will incur a penalty of 2 marks per day unless prior approval is given.

Academic Honesty

Plagiarism or copying from peers will result in **zero marks** for the assignment. You may discuss concepts, but **write your own steps and reasoning**.

Submission Format Example

- **Page 1:** Name, Roll No., Section, Assignment Title
- **Pages 2–n:** Answers with clearly labeled question numbers
- Highlight important steps or box final answers
- If coding: attach output screenshots and explain in 2–3 lines