Math: Gradient Overkill

Problem Story

For her final task, Ivan asked Alice to help him with implementing "Green Line" ML task. In particular, he asked to explore mathematical foundations of Gradient Descent algorithm. Help Alice to complete her assignment!

Problem

We are trying to train an exceptionally complicated model (1D linear regression, aka a point) on an exceptionally complicated dataset (1 1D point located at $x = 5$), and we need your help...

Essentially, for such small dimensionality, the problem is reduced to searching for the point by the linear regression model. To define how close we are, we use a so-called loss function. For this problem, it is defined as the square of the distance between the current point and the point $x = 5$:

$$L(x) = (x - 5)^2$$

You can verify that the value of the loss function is minimal and equal to zero at $x = 5$.

The model starts from an arbitrary point x_0 and attempts to reach a point $x = 5$. For every step, it checks whether going to the right increases the value of the loss function. It does so by calculating a *slope* or *derivative* of the loss function L *with respect to* its *argument* x (or, for step t, x_t), denoted as $\frac{dL}{dx_t}$. Intuitively, one should move to a position of smaller value of L. When going to the right increases the loss function (which means the derivative is positive), it makes sense to go to the left, and vice versa. To calculate how large the step should be, we take the derivative and multiply it by a constant α. The choice of α is usually up to you (and, spoiler alert, you will explore how to choose this constant). For now, assume α to be positive.

More formally, the transition from step t to $t + 1$ is given by the formula:

$$x_{t+1} = x_t - \alpha \cdot \frac{dL}{dx_t}$$

The derivative is given as

$$\frac{dL}{dx_t} = 2 \cdot (x_t - 5)$$

Take your time to verify that, if the derivative is positive, the step is made to the left and vice versa.

If you are more familiar with Machine Learning, you may recognize the procedure to be known as gradient descent, and α as the learning rate.

Your task is the following:

{1} Run the gradient descent starting from $x_0 = 0$, with $\alpha = 0.3$. You may stop if the value of the loss function L at step x_t is at most 1. For each step, show the following: $x_t, L(x_t), \frac{dL}{dx_t}, x_{t+1}$.

{2} Explore how gradient descent proceeds for different values of α. You will notice that the model will get closer to 5 for some values of α and will not for others. Describe for which values of α the model converges (or doesn't) to 5. Try to identify the range of values of α for which gradient descent

converges to 5. Make sure to analyze the boundary points of the range you find. Your explanation should be as full as possible and include analysis for all possible cases.

Hint: in your search for the answer, you might find the properties of geometric series and limits useful.

Format

For {1}, provide a table with columns: t, x_t, $L(x_t)$, $\frac{dL}{dx_t}$, x_{t+1}.

For {2}, give a mathematical explanation. We don't require a formal mathematical proof, but to get full points, you should provide a valid mathematical explanation on deriving interval for α, not just observations.

Include relevant details, but avoid unnecessary or lengthy elaboration (if you don't provide formal proof, try to keep solution limited to half of A4 page or less). That will help us grade your solution faster and more accurately. You may provide your solution in Latvian, Russian, or English.

Grading

- 25% for correct {1}
- 75% for correct {2}

Partial points are obtainable on both parts, for partially correct answers.

Further explanation

If you wish to find out more about gradient descent, and its use in Machine Learning, we reccomend:
- short video by Visually Explained: https://www.youtube.com/watch?v=qg4PchTECck
- longer explanation by 3b1b (part of his Neural Networks series): https://www.youtube.com/watch?v=IHZwWFHWa-w