Assignment: Gradient Descent for Unconstrained Minimization

Instructions

You will use **Python or R** to implement gradient descent to minimize the following unconstrained functions. For each problem:

- Implement gradient descent using a **fixed step size**.
- Stop when the gradient norm satisfies $\|\nabla f(x)\| < \varepsilon$, or after a maximum number of iterations.
- Plot the loss vs. iterations and, where applicable, a 2D or 3D visualization of the descent path.

General Parameters

For each problem, you will be provided:

- The objective function $f(x)$
- An initial starting point x_0
- Step size η
- Threshold ε

Problem Set

Problem 1: Linear Regression Loss

Function:

$$f(\beta) = \frac{1}{2n} \sum_{i=1}^{n} \left(y_i - \mathbf{x}_i^\top \beta\right)^2$$

Details:

- Generate $n = 100$ paired (x, y) where $x \sim N(1, 2)$ and $y \sim N(2+3x, 5)$ data ($n = 100$, $p = 2$)
- Initial point: $\beta_0 = [0, 0]$
- Step size: $\eta = 0.01$
- Threshold: $\varepsilon = 10^{-6}$

Problem 2: Logistic Regression Negative Log-Likelihood

Function:

$$f(\beta) = \sum_{i=1}^{n} \log\left(1 + \exp(-y_i \mathbf{x}_i^\top \beta)\right)$$

Gradient:

$$\nabla f(\beta) = -\sum_{i=1}^{n} y_i \mathbf{x}_i \left(1 - \sigma(y_i \mathbf{x}_i^\top \beta)\right) \quad \text{where } \sigma(z) = \frac{1}{1 + e^{-z}}$$

Details:

- Binary classification problem (data provided here in the link `https://docs.google.com/spreadsheets/d/13CmIStaYtiQqR_dhBPrkHJINvVln9cepHypNinVQT3c/edit?gid=0#gid=0`)
- Initial point: $\beta_0 = [0, 0]$
- Step size: $\eta = 0.05$
- Threshold: $\varepsilon = 10^{-5}$

Problem 3: Quadratic Convex Function

Function:

$$f(x) = x^\top A x + b^\top x \quad \text{with } A = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}, \quad b = \begin{bmatrix} -4 \\ -8 \end{bmatrix}$$

Details:

- Initial point: $x_0 = [1, 1]$
- Step size: $\eta = 0.1$
- Threshold: $\varepsilon = 10^{-6}$

Problem 4: Normal Distribution MLE (Negative Log-Likelihood)

Function:

$$f(\mu, \sigma) = \sum_{i=1}^{n} \log(\sigma) + \frac{(x_i - \mu)^2}{2\sigma^2}$$

Note: Treat $\mu, \sigma > 0$ as unconstrained variables for gradient descent.
Details:

- Data: $x_1, \ldots, x_n$ provided here:`https://docs.google.com/spreadsheets/d/13CmIStaYtiQqR_dhBPrkHJINvVln9cepHypNinVQT3c/edit?gid=2023320122#gid=2023320122`
- Initial point: $\mu_0 = 0$, $\sigma_0 = 1$
- Step size: $\eta = 0.01$
- Threshold: $\varepsilon = 10^{-5}$

Problem 5: Rosenbrock Function (Non-Convex)

Function:

$$f(x, y) = (1 - x)^2 + 100(y - x^2)^2$$

Gradient:

$$\nabla f(x, y) = \begin{bmatrix} -2(1 - x) - 400x(y - x^2) \\ 200(y - x^2) \end{bmatrix}$$

Details:

- Initial point: $[x_0, y_0] = [-1, 1]$
- Step size: $\eta = 0.001$
- Threshold: $\varepsilon = 10^{-6}$

Note: This is a challenging non-convex function with a narrow curved valley.

Submission

Please prepare and submit

- Your code (in Python or R)
- A short report with:
  - Plots of convergence
  - Final solution and number of iterations
  - Any observations or difficulties encountered