Gradient Descent with 1 dimension:

Minimize loss function $L(x) = 2x^2 + 4x + 5$ with learning rate $= 0.1$

Step 1: Calculate derivative function

$$\nabla L(x) = 4x + 4$$

Step 2: Find value of x when $L(x)$ is minimized

$$\nabla L(x) = 0$$

$$\Leftrightarrow \quad 4x + 4 = 0$$

$$\Leftrightarrow \quad x = -1$$

So when $x = -1$, the loss function is minimized

Step 3: Iteratively minimize loss function

1st round:

- Initial guess: $x_0 = 2$
- $x_1 := x_0 - \eta \nabla L(x_0)$
  $= 2 - 0.1 \cdot 12$
  $= 0.8$

2nd round:

- $x_2 := x_1 - \eta \nabla L(x_1)$
  $= 0.8 - 0.1 \cdot (7.2)$
  $= 0.08$

3rd round:

- $x_3 := x_2 - \eta \nabla L(x_2)$
  $= 0.08 - 0.1 \cdot (4.32)$
  $= -0.35$

4th round:

- $x_4 := x_3 - \eta \nabla L(x_3)$
  $= -0.35 - 0.1 \cdot (2.59)$
  $= -0.61$

5th round:

- $x_5 := x_4 - \eta \nabla L(x_4)$
  $= -0.61 - 0.1 \cdot (1.56)$
  $= -0.77$

6th round:

- $x_6 := x_5 - \eta \nabla L(x_5)$
  $= -0.77 - 0.1 \cdot (0.93)$
  $= -0.86$

7th round:

- $x_7 := x_6 - \eta \nabla L(x_6)$
  $= -0.86 - 0.1 \cdot (0.56)$
  $= -0.92$
  ≈ -1

Gradient Descent Example with 2 dimensions

Minimize this loss function $L(x,y) = 0.5x^2 + 2x + y^2 + y + 3$, $\eta = 0.2$

Step 1: Calculate derivative function

$$\nabla_x L(x,y) = x + 2$$

$$\nabla_y L(x,y) = 2y + 1$$

Step 2: Find value of x, y when $L(x,y)$ is minimized

$$\nabla_x L(x,y) = x + 2 = 0$$

$$\Rightarrow x = -2$$

$$\nabla_y L(x,y) = 2y + 1 = 0$$

$$\Rightarrow y = -1/2$$

Step 3: Iteratively minimize loss function

- Initial guess: $x_0 = 0$, $y_0 = 0$
- 1st round:
  - $x_1 := x_0 - \eta \cdot \nabla_x L$
    $= 0 - 0.2 \cdot (2)$
    $= -0.4$
  - $y_1 := y_0 - \eta \nabla_y L$
    $= 0 - 0.2 \cdot (1)$
    $= -0.2$
- 2nd round:
  - $x_2 := x_1 - \eta \cdot \nabla_x L$
    $= -0.4 - 0.2 \cdot (1.6)$
    $= -0.72$
  - $y_2 := y_1 - \eta \nabla_y L$
    $= -0.2 - 0.2 (0.6)$
    $= -0.32$
- 3rd round:
  - $x_3 := x_2 - \eta \cdot \nabla_x L$
    $= -0.72 - 0.2 \cdot (1.28)$
    $= -0.976$
  - $y_3 := y_2 - \eta \nabla_y L$
    $= -0.32 - 0.2 (0.36)$
    $= -0.392$
- 4th round:
  - $x_4 := x_3 - \eta \cdot \nabla_x L$
    $= -0.976 - 0.2 \cdot (1.024)$
    $= -1.1808$
  - $y_4 := y_3 - \eta \cdot \nabla_y L$
    $= -0.392 - 0.2 \cdot (0.216)$
    $= -0.435$
- 5th round:
  - $x_5 := x_4 - \eta \cdot \nabla_x L$
    $= -1.1808 - 0.2 (0.819)$
    $= -1.345$
  - $y_5 := y_4 - \eta \cdot \nabla_y L$
    $= -0.435 - 0.2 \cdot (0.13)$
    $= -0.461$

- 6^{th} round:

$$\begin{aligned} x_6 &:= x_5 - \eta \nabla_x L \\ &= -1.345 - 0.2 \cdot (0.655) \\ &= -1.476 \end{aligned}$$

$$\begin{aligned} y_6 &:= y_5 - \eta \nabla_y L \\ &= -0.461 - 0.2 (0.078) \\ &= -0.477 \end{aligned}$$

- 7^{th} round:

$$\begin{aligned} x_7 &:= x_6 - \eta \nabla_x L \\ &= -1.476 - 0.2 \cdot (0.524) \\ &= -1.58 \end{aligned}$$

$$\begin{aligned} y_7 &:= y_6 - \eta \nabla_y L \\ &= -0.477 - 0.2 \cdot (0.046) \\ &= -0.486 \end{aligned}$$

- 8^{th} round:

$$\begin{aligned} x_8 &:= x_7 - \eta \nabla_x L \\ &= -1.58 - 0.2 \cdot (0.42) \\ &= -1.664 \end{aligned}$$

$$\begin{aligned} y_8 &:= y_7 - \eta \nabla_y L \\ &= -0.486 - 0.2 \cdot (0.028) \\ &= -0.492 \end{aligned}$$

- 9^{th} round:

$$\begin{aligned} x_9 &:= x_8 - \eta \nabla_x L \\ &= -1.664 - 0.2 \cdot (0.336) \\ &= -1.731 \end{aligned}$$

$$\begin{aligned} y_9 &:= y_8 - \eta \nabla_y L \\ &= -0.492 - 0.2 \cdot (0.016) \\ &= -0.495 \end{aligned}$$

- 10^{th} round:

$$\begin{aligned} x_{10} &:= x_9 - \eta \nabla_x L \\ &= -1.731 - 0.2 \cdot (0.269) \\ &= -1.785 \end{aligned}$$

$$\begin{aligned} y_{10} &:= y_9 - \eta \nabla_y L \\ &= -0.495 - 0.2 \cdot (0.01) \\ &= -0.497 \end{aligned}$$

- 11^{th} round:

$$\begin{aligned} x_{11} &:= x_{10} - \eta \nabla_x L \\ &= -1.785 - 0.2 \cdot (0.215) \\ &= -1.828 \end{aligned}$$

$$\begin{aligned} y_{11} &:= y_{10} - \eta \nabla_y L \\ &= -0.497 - 0.2 \cdot (0.006) \\ &= -0.498 \end{aligned}$$

- 12^{th} round:

$$\begin{aligned} x_{12} &:= x_{11} - \eta \nabla_x L \\ &= -1.82 - 0.2 \cdot (0.172) \\ &= -1.862 \end{aligned}$$

$$\begin{aligned} y_{12} &:= y_{11} - \eta \nabla_y L \\ &= -0.498 - 0.2 \cdot (0.004) \\ &= -0.497 \end{aligned}$$

- 13^{th} round:

$$\begin{aligned} x_{13} &:= x_{12} - \eta \nabla_x L \\ &= -1.862 - 0.2 \cdot (0.138) \\ &= -1.89 \\ &\approx -2 \end{aligned}$$

$$\begin{aligned} y_{13} &:= y_{12} - \eta \nabla_y L \\ &= -0.497 - 0.2 \cdot (0.006) \\ &= -0.498 \\ &\approx -1/2 \end{aligned}$$

How to prove a function is convex function

Definition of convex: $f(\theta x + (1-\theta)y) \leqslant \theta f(x) + (1-\theta) f(y)$

where $0 \leqslant \theta \leqslant 1$; $x, y \in \text{domain}(f)$

- Prove $f(x,y) = 0.5x^2 + 2x + y^2 + y + 3$

Step 1: Compute first partial derivatives

$$\nabla_x f = x + 2$$

$$\nabla_y f = 2y + 1$$

Step 2: Compute second partial derivative

$$\nabla_{x^2} f = 1 \qquad \nabla_{x,y} f = \nabla_{y,x} f = 0$$

$$\nabla_{y^2} f = 2$$

Step 3: Form Hessian matrix

$$\begin{bmatrix} \nabla_{x^2} f & \nabla_{x,y} f \\ \nabla_{y,x} f & \nabla_{y^2} f \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

Step 4: Check if Hessian matrix is positive semi-definite

Here the matrix is diagonal with non negative diagonal entries

$\Rightarrow$ Hessian matrix is positive semi-definite

- Prove $f(x) = 2x^2 + 4x + 5$ is convex

Step 1: Find first derivative

$$f'(x) = 4x + 4$$

Step 2: Find second derivative

$$f''(x) = 4$$

Step 3: Analyze second derivative

Since $f''(x) = 4 \geqslant 0$ for all x in domain of f,

f is a convex function

- Prove this function $f(x) = \left(\max(0,x) - \frac{1}{2}\right)^2$ is convex

Step 1: Find first derivative

$$f'(x) = \begin{cases} 2x - 1 & \text{if } x > 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Step 2: Find second derivative

$$f''(x) = \begin{cases} 2 & \text{if } x > 0 \\ 0 & \text{if } x < 0 \end{cases}$$

Step 3: Analyze the second derivative

Ignore the discontinuity at 0. We can see that $f''(x) \geq 0 \quad \forall x \in dom(f)$

So $f(x)$ is convex everywhere except when $x = 0$