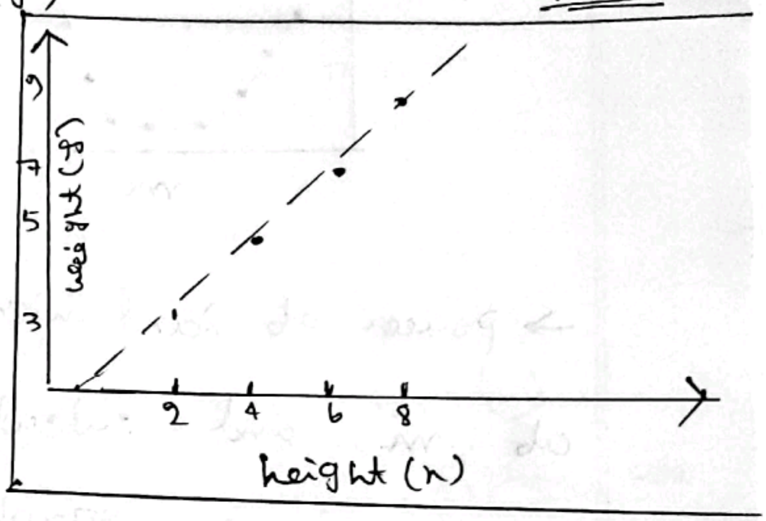
gradient descent

Dt: 30/07/25

Assume a dataset,

height(x)	weight (y)
2	3
4	5
6	7
8	9

Fig 1



<u>Problem 01</u>: The relationship between x and y can be defined by $y = mx$. So, you can predict y (weight) by plugging in x (height) in the equation.

For example, y (weight) $= m \times 7$

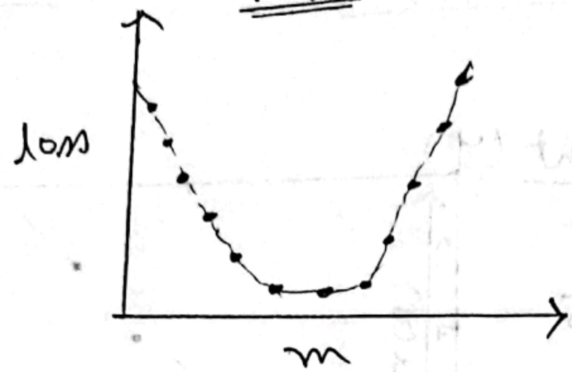
Question is, how can we find m ???

→ Using Least Square Method [STA201]

→ Using Gradient Descent [CSE422/CSE437]

(*) GD is an opt. algorithm, and have plenty of usage: minimizing error (loss function)

Fig 2



Remember: Our target is to find optimal value of "m", so that loss is minimized

→ Power of Randomization: Randomly pick values of m, and calculate error (loss function). Plot them in a graph, and easily get the optimal solution.

⊛ This is how nature works. It always select the best candidate (Survival of the fittest).

But too randomness will not work. What if optimal solution is within the interval [2, 5], but you are randomly trialing within [7, 10]?

→ $\boxed{\text{Gradient descent solves this problem}}$

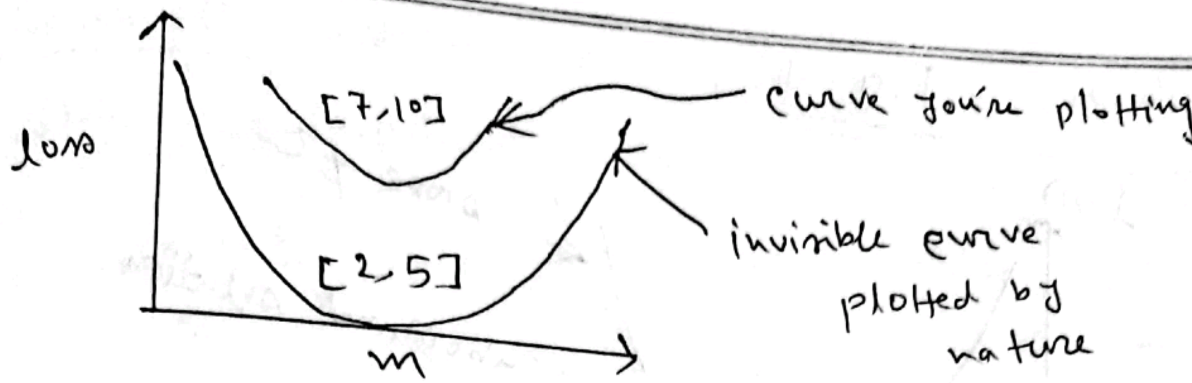


Fig 3

How GD solves the problem?

1. We know the curve of loss vs m is ∪

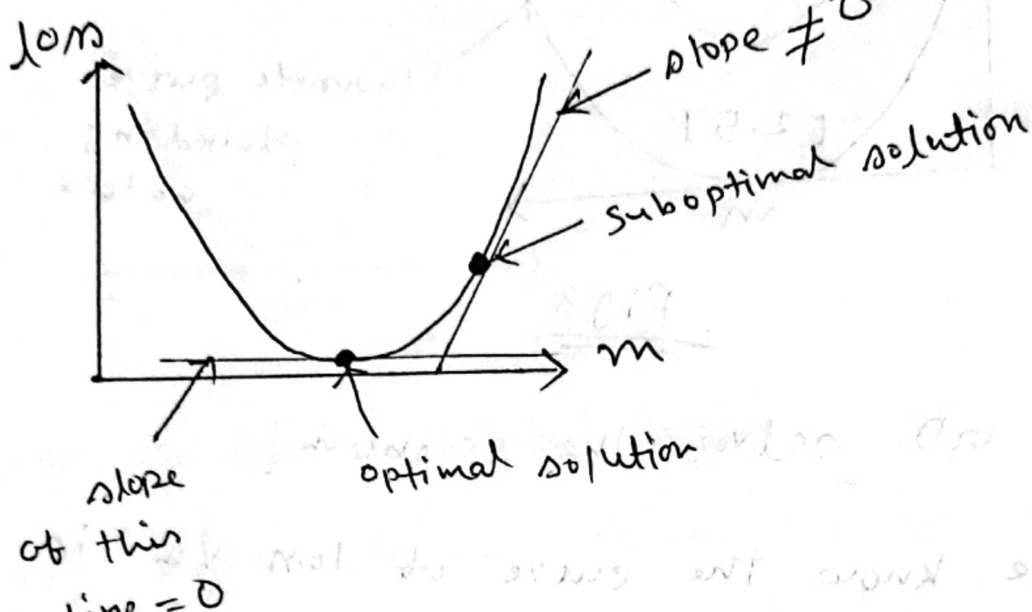
2. If we can somehow reach at the bottom, we get our optimal solution.

3. Slope of a horizontal line is 0 (as $\Delta y = 0$).

so, when we reach our optimal solution:

Derivative of the loss function with respective to the parameter (m) is 0. Or, $\frac{dl}{dm} = 0$

Fig 4



Hence, if $\frac{dL}{dm} \neq 0$, we keep updating m.

But how much do we update m?

Update rule: $m = m - \frac{dL}{dm} \times \alpha$, where,

α = learning rate

$\boxed{\text{Simulation}}$ If, $m = 2.5$ and $\alpha = 0.1$

new $m = 2.5 - (2.5 \times 0.1)$

$= 2.25$ (updated by 0.25)

If, $m = 0.15$,

new $m = 0.15 - (0.15 \times 0.1)$

$= 0.135$ (updated by 0.015)

Exercise 01: Perform one epoch and find the value ~~of~~ (updated value) of m. $\boxed{\alpha = 0.01}$

Use SSR as loss function.

height	weight
2	3
4	5
6	7
8	9

Solⁿ: $y = mx$. Let, $m = 0.1$

Step 01 - Calculate loss function.

$$\boxed{\begin{aligned} y_1 &= 0.1 x_1 = 0.1 \times 2 = 0.2 \\ \text{loss} &= |(3 - 0.2)| = 2.8 \end{aligned}}$$

not necessary in math. Just to make ur understand

Squared Residual for T1:

$$(3 - m \times 2)^2 = (3 - 2m)^2$$

Squared Residual for T2:

$$(5 - m \times 4)^2 = (5 - 4m)^2$$

for T3: $(7 - 6m)^2$

for T4: $(9 - 8m)^2$

$\therefore L$ = Sum of squared residuals (SSR)

$= T1_{sr} + T2_{sr} + T3_{sr} + T4_{sr}$

$= (3-2m)^2 + (5-4m)^2 + (7-6m)^2 + (9-8m)^2$

Step 02: Calculate $\frac{dL}{dm}$

$$\frac{dL}{dm} = \frac{d}{dm}\left[(3-2m)^2 + (5-4m)^2 + (7-6m)^2 + (9-8m)^2\right]$$

$$= \frac{d}{dm}(3-2m)^2 + \frac{d}{dm}(5-4m)^2 + \frac{d}{dm}(7-6m)^2 + \frac{d}{dm}(9-8m)^2$$

$$= 2(3-2m).-2 + 2(5-4m).-4 + 2(7-6m).-6 + 2(9-8m).-8$$

$$= -11.2 - 36.8 - 76.8 - 131.2 = -233.6$$

Step 03: Calculate updated m.

$$m = m - \frac{dL}{dm} \times \alpha$$

$$= 0.1 - (-233.6 \times 0.01)$$

$$= 2.436 \text{ (Ans)}$$

$$y_1 = 2.436\, x_1 = 2.436 \times 2 = 4.872$$

$$loss = |(3 - 4.872)| = 1.87 < 2.8$$

Previous loss

Example 2: The following data can be modeled by: $y = m_1 x_1 + m_2 x_2 + c$. Find the first updated value of m_1, m_2, and c using Gradient Descent. Assume, $m_1 = 0.1$, $m_2 = 0.3$, $c = 1$, $\alpha = 0.00025$. Use SSR as loss f.

age	yexp	salary (k)
28	4	40
31	7	44
33	8	45

Solⁿ: Step 01: calculate SSR.

$T1_{sr} = (40 - 28m_1 - 4m_2 - c)^2$

$T2_{sr} = (44 - 31m_1 - 7m_2 - c)^2$

$T3_{sr} = (45 - 33m_1 - 8m_2 - c)^2$

$$SSR = l = (40-28m_1-4m_2-c)^2 + (44-31m_1-7m_2-c)^2 + (45-33m_1-8m_2-c)^2$$

Step 02: Calculate $\frac{dl}{dm_1}$, $\frac{dl}{dm_2}$, $\frac{dl}{c}$

$$\frac{dl}{dm_1} = \frac{d}{dm_1}\left[(40-28m_1-4m_2-c)^2\right]$$
$$+ \frac{d}{dm_1}\left[(44-31m_1-7m_2-c)^2\right]$$
$$+ \frac{d}{dm_1}\left[(45-33m_1-8m_2-c)^2\right]$$

$$\left.\begin{aligned} &= 2(40-28m_1-4m_2-c).(-28) \\ &+ 2(44-31m_1-7m_2-c).(-31) \\ &+ 2(45-33m_1-8m_2-c).(-33) \end{aligned}\right\} \text{Partial derivative}$$

$$= -1960 - 2343.6 - 2527.8 = -6831.4$$

Updating m_1: $m_1' = m_1 - \left(\frac{dl}{dm_1}\alpha\right)$

$$= 0.1 - (-6831.4 \times 0.00025)$$

$$= \boxed{1.80}$$

$$\frac{dL}{dm_2} = 2(40 - 28m_1 - 4m_2 - c).(-4)$$
$$+ 2(44 - 31m_1 - 7m_2 - c).(\cancel{8}7)$$
$$+ 2(45 - 33m_1 - 8m_2 - c).(-8)$$

$$= -280 - 529.2 - 612.8 = -1422$$

$$m_2' = m_2 - \left(\frac{dL}{dm_2}\alpha\right) = 0.3 - (-1422 \times 0.00025)$$
$$= \boxed{0.655}$$

$$\frac{dL}{dc} = 2(40 - 28m_1 - 4m_2 - c).-1$$
$$+ 2(44 - 31m_1 - 7m_2 - c).-1$$
$$+ 2(45 - 33m_1 - 8m_2 - c).-1$$

$$= \cancel{1.05}\ -222.2$$

$$c' = c - \left(\frac{dL}{dc}\alpha\right) = 1 - (\cancel{1.05}\ -222.2 \times 0.00025)$$
$$= \boxed{1.05}$$

So, updated $(m_1, m_2, c) = (1.80, 0.655, 1.05)$

(Ans)