LR with Gradient Descent

$y = mx + b$

② 'x' → I/P data

from 'x' I try to make a guess

$X \rightarrow$ ~~f~~ (guess)

$X \xrightarrow{\text{M/c}} \text{guess}$ → [Y] → original value

Error = y − guess → Cost fnc.

$$\text{COST} = \sum_{i=1}^{N} \left(y_i - \text{guess}_i \right)^2$$

↳ Total error for the particular model; ~~root~~

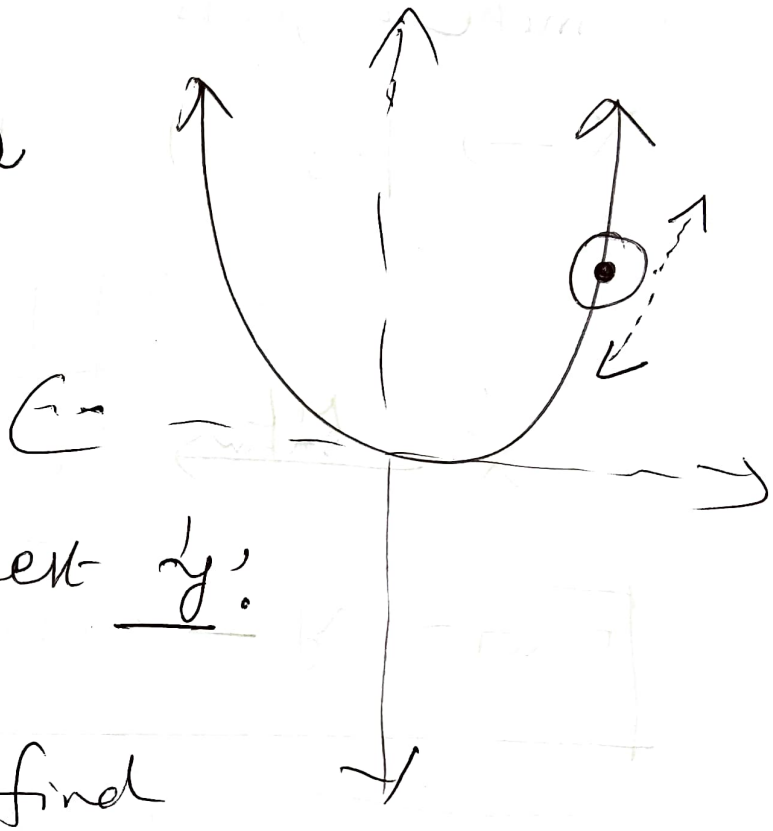
where i is count values of 'm' & 'b' denote the particular line.

Goal: Minimize the loss or cost fn(). (or) I want the smallest error

Minimizing a fn():

eg: $y = x^2$

(or) $f(x) = x^2$



i.e.

Find a 'x' value that gives the lowest 'y'.

How would I find the minimum)

(1) Direction of movement from given point (ascent or descent)

(2) Step-Size = big or small.

2202-03-04

(c)

Using GD to find the optimal value of the intercept

Sol: In this case the RV=0

$$\text{SS Residuals} = (\text{Observed} - \text{Predicted})^2$$

$$\text{SSR1} = (1.4 - (\text{intercept} + 0.64 * \text{height}))^2$$

$$= (1.4 - (\text{intercept} * (0.64 + 0.5)^2))^2$$

Now we can plug in any value for intercept and get a new predicted height.

$$\text{Total SSR} = (1.4 - (\text{intercept} + 0.64 * \text{height}^{0.5}))^2$$

$$+ (1.9 - (\text{intercept} + 0.64 * 2.3))^2$$

$$+ (3.2 - (\text{intercept} + 0.64 * 2.9))^2$$

This gives an eqn. for the curve

Step-01: Take the derivative of SSR's w.r.t. the 'Intercept Term'.

$$\frac{d}{d\,\text{intercept}}\left(\text{total SSR's}\right) = d(SSR1) + d(SSR2) + d(SSR3)$$

Apply chain rule

$$\frac{d}{d\,\text{intercept}}(SSR1) = \frac{d}{d\,\text{intercept}}\left(1.4 + (C_1)\,\text{intercept} - 0.64 * 05\right)^2$$

$$= -2(1.4 - (\text{intercept} + 0.64 * 05))$$

$$\therefore \frac{d}{d\,\text{intercept}}(SSR1) = -2(1.4 - \text{intercept} + (0.64 * 05))$$

Step: Take the derivative of SSR's w.r.t. the intercept term.

$$\text{Residual} = (\text{Observed} - \text{predicted})^2$$

$$\frac{d}{d\text{intercept}} (\text{Total SSR's})$$

$$= d(SSR1) + d(SSR2) + d(SSR3)$$

$$\frac{d}{d\text{intercept}} (SSR1) = \frac{d}{d\text{intercept}} \left(1.4 + (-1)\,\text{intercept} - 0.64 + 0.5\right)$$

$$= -2\left(1.4 - (\text{intercept} + 0.64 \times 0.5)\right)$$

$$\frac{d}{d\text{intercept}} (SSR2) = -2\left(1.9 - (\text{intrcept} + 0.64 \times 2.3)\right)$$

$$(SSR3) = -2\left(3.2 - (\text{intercept} + 0.64 \times 2.9)\right)$$

Now $\frac{d}{d\text{intercept}} (\text{Total SSRS}) = d(SSR1) + d(SSR2) + d(SSR3)$

Now with this derivative value, GD will use it to find where the SSR is the lowest.

Note: If we were using least Squares to solve for the optimal value of the intercept, we would simply find where the slope of the curve = 0.

In contrast, GD finds the min. value by taking steps from an initial guess until it reaches the best value.

This makes GD very useful when it is not possible to solve for where the derivative = 0.

Our initial value of intercept = 0.

$$\therefore \frac{d}{d\text{intercept}}(SSR'1) =$$

$$-2(1.4 - (0 + 0.64 * \frac{0.32}{1.472})) = -2.16$$

$$+ -2(1.9 - (0 + 0.64 * 2.3)) = -0.856$$

$$+ -2(3.2 - (0 + 0.64 * \frac{1.856}{2.9})) = -2.688$$

$$= -5.704$$

$$\Rightarrow \frac{d}{d\,\text{intercept}}(SSR's) = -5.7$$

⇒ When intercept = 0; slope of the curve = −5.7 ⇒

Note: The closer we set to the optimal value for the intercept, the closer the slope of the curve sets to zero.

⇒ we need to take "Baby steps" since we are close to the optimal value.

And we should take big steps when we are far from optimal value.

⇒ Size of the step is related to the slope; as it tells whether we should take a baby step or big step

Note: We should ensure that the big step is not too big.

Gradient descent determines the step-size by multiplying the slope by a small no. called the "Learning Rate".

$$\text{Step-Size} = -5.7 * 0.1 = -0.57$$

With step-size we can calculate a new intercept

$$\text{New Intercept} = \text{old Intercept} - \text{Step-Size}$$

Now the

New-intercept
$= 0 - (-0.57)$
$= 0.57$

w.r.t. the original line & original intercept, we see that the "residuals shrank" when the intercept $= 0.57$.

To take another step, we go back to the derivative & plug in the new intercept $(= 0.57)$

$$\frac{d}{d\text{intercept}}(SSR) =$$
$$-2(1.4 - (0.57 + 0.64 \times 0.5))$$
$$+ \quad -2(1.9 - (0.57 + 0.64 \times 2.3))$$
$$+ \quad -2(3.2 - (0.57 + 0.64 \times 2.9))$$
$$= -2.3 \quad (= \text{new slope of the curve})$$

Now calculate the step-size:

Step-size = slope * learning rate

$= -2.3 * 0.1$

Step-size $= -0.23$

∴ New intercept $= 0.57 - (-2.3)$

$= 0.57 + 0.23$

New-intercept $= 0.8$

Now we can compare the residuals when the intercept $= 0.57$ & new-intercept $= 0.8$;

Overall the SSR is getting smaller.

1st step-size = 0 to 0.57 (relatively large)

2nd " = 0.57 to 0.8

$$\frac{d}{d\text{intercept}}(SSR/5) = -2.288 \approx -2.3 \; (= \hat{2}.3)$$

Now calculate the derivative at the new intercept:

$$\frac{d}{d\text{intercept}}(\text{Sum of Squared residuals})$$

$$= (-2)(1.4 - (0.8 + 0.64 * 0.5))$$

$$+ (-2)(1.9 - (0.8 + 0.64 * 2.3))$$

$$+ (-2)(3.2 - (0.8 + 0.64 * 2.9))$$

$$= -0.9 \rightarrow \text{New-intercept}$$

$$\text{Step-size} = -0.9 * 0.1 = -0.09$$

$$\text{New-intercept} = 0.8 - (-0.09)$$

$$= 0.89$$

Now we increase the intercept from 0.8 to 0.89.

Then we take another step & New-intercept = 0.92

Another step: New-intercept = 0.94

" : " = 0.95

Notice: Each step gets smaller & smaller the closer we get to the bottom of the curve.

After 6 steps, the GD estimate for the intercept is 0.95

Note: The least squares estimate for the intercept is also 0.95.

$$\frac{d}{d\text{intercept}}(SSR/5) = -2.288 \approx -2.3 \left(= \hat{2}\cdot\bar{3}\right)$$

Now calculate the derivative at the new intercept:

$$\frac{d}{d\text{intercept}}(\text{Sum of Squared residuals})$$

$$= (-2)(1.4 - (0.8 + 0.64 * 0.5))$$

$$+ (-2)(1.9 - (0.8 + 0.64 * 2.3))$$

$$+ (-2)(3.2 - (0.8 + 0.64 * 2.9))$$

$$= -0.9 \rightarrow \text{New-intercept}$$

$$\text{Step-size} = -0.9 * 0.1 = -0.09$$

$$\text{New-intercept} = 0.8 - (-0.09)$$

$$= 0.89$$

Now we increase the intercept from 0.8 to 0.89.

Then we take another step & new-intercept = 0.92

Another step: new-intercept = 0.94

= 0.95

Notice: Each step gets smaller & smaller the closer we get to the bottom of the curve.

After 6 steps, the GD estimate for the intercept is 0.95

Note: The least squares estimate for the intercept is also 0.95.

Now we assume that GD has done its job.

But without comparing its solⁿ to a "Gold std.", how does GD know to stop taking steps?

GD Termination Steps

(1) when Step-size is very close to 0. (= when slope is very close to zero)

(2) In practice, the minimum Step-size = 0.001 (or) smaller

Step-size = slope * Learning Rate

If slope = 0.009;

then

step-size = 0.009 * 0.1 (learning rate)

= 0.0009

(< 0.001)

Now the GD stops.

(3) GD also includes a limit on the no. of steps it will take before giving up.

In practice, the max. no. of steps = 1000, or greater

Even if step-size is large, if the max. no. of steps > max. steps, GD will stop.

Summary:

(1) Use of SSR as Loss fnc), to evaluate how well a line fits the data

(2) we take the derivative of SSR's i.e. the derivative of the Loss Function.

$$\frac{d}{d\,\text{intercept}}(SSR)$$

(3) Pick a RV for intercept =0 in our case.

(4) we calculate the derivative when intercept =0. (O/p = slope of the curve)

(5) Plug the resultant slope value into Step-Size calculation.

Step-Size = Slope * Learning Rate

(6) Calculate New Intercept Value as:

$$\text{New Intercept} = \text{old-intercept} - \text{Step-Size}$$

(7) Plug the New-intercept into the derivative & repeat the steps until the size goes close to 0

$$\text{step-size} = \text{slope} * \text{learning Rate}$$

$$\text{New-Intercept} = \text{old-intercept} - \text{step-size.}$$

We understood how GD can estimate the Intercept;

let us understand how to estimate the Intercept & Slope.

(1) Use SSR as before as the Loss Function.

(2) We have a 3-D graph:

x-axis = Slope-values

y-axis = SSR

z-axis = Intercept values.

(3) We need to find the values for the intercept & Slope that gives us the min. SSR value.

(4) SSR =

$$(1.4 - (\text{intercept} + \text{slope} * 0.5))^2$$

$$+ (1.9 - (\quad " \quad + " \quad * 2.3))^2$$

$$+ (3.2 - (\quad " \quad + " \quad * 2.9))^2$$

Take the derivative of this fnc.

" " " w.r.t. Intercept

$$\frac{d}{d\,\text{intercept}} (SSR) \text{ w.r.t. Intercept}$$

+ we also take the derivative of loss fnc (SSR) w.r.t. "slope"

(5) Start taking the derivative w.r.t. the intercept. (Now slope is constant)

$$\frac{d}{d\text{intercept}}(SSR) =$$

$$(-2)(1.4 - (\text{intercept} + \text{slope} * 0.5))$$
$$+ (-2)(1.9 - (\text{intercept} + \text{slope} * 2.3))$$
$$+ (-2)(3.2 - (\text{intercept} + \text{slope} * 2.9))$$

Note:
Derivative of Constant = Zero

(6) Take derivative of Loss Fnc (=SSR) w.r.t. the slope:

→ intercept = constant

$$\frac{d}{d\text{slope}}(SSR) = -0.5 * 2(1.4 - (\text{intercept} + (\text{slope} * 0.5)))$$
$$+ (-2) * 2.3(1.9 - (\text{intercept} + (\text{slope} * 2.3)))$$

$+(-2)*(2.9)\left(3.2-\left(\text{intercept}+\text{slope}*2.9\right)\right)$

Note: When u have 2 or more derivatives of the same fnc, they are called a "Gradient".

We Use this gradient to descend to lowest point in the Loss Fnc, i.e. the SSR.

Step 1: (i) Pick a random no. for the intercept. (=0)

(ii) Pick a Random no. for the slope. (Pick slope = 1)

Starting point :

Intercept = 0

Slope = 1.

left the Learning Rate = 0.01

Plug in the values

$$\Rightarrow \frac{d}{d\text{intercept}} (SSR) = -1.6 \quad \&$$

$$\frac{d}{d\text{slope}} (SSR) = -0.8$$

$$\text{Step-Size}_{\text{intercept}} = -1.6 * 0.01$$

$$\text{Step-Size}_{\text{slope}} = -0.8 * 0.01$$

Note: The larger learning rate of the former. eg. (= 0.1) doesn't work this time.

Even after a bunch of steps, GD doesn't arrive at the correct answer.

⇒ GD is very sensitive to "Learning Rate".

Note: In practice, a reasonable learning Rate can be determined automatically by starting large & getting smaller with each step.

$$= -2(1.4 - (0.57 \underbrace{0.89}_{+0.32}))$$

$$+ -2(1.9 - (\underbrace{0.57 + 1.472}_{2.042}))$$

$$+ -2(3.2 - (\underbrace{0.57 + (1.856)}_{2.426}))$$

$$= (-2)(0.51) + (-2)(-0.142) + (-2)(0.770)$$

$$= -1.02 + 0.284 - 1.542$$

$$= -2.542 + 0.284$$

$$= -2.288 \approx 2.3$$