GD

1) predicted Height = Intercept + slope * weight

Assume: slope = 0.64

GD: Finds optimal value for intercept.

To start with find a R.V. for intercept
(= 0)

$\therefore$ PH = ~~0~~ Intercept + 0.64 * weight

How well the line fits the data
with SSRs. (= loss fncl)

Start by calculating the newdeal.

$\boxed{\text{Intercept} = 0;}$

Data pts

Wt	Height	PH
0.5	1.4	0.32

predicted Height = point on the line

Weight	Height	PH	Residual
0.5	1.4	0.32	1.1
2.3	1.9	~~0.4~~	0.4
~~2.2~~ 2.9	3.2		1.3

$$SSR = (1.1)^2 + (0.4)^2 + (1.3)^2 = 3.1$$

SSR Vs Intercept

Intercept = 0, 0.25, 0.5, ...

Find lowest values

We can do this until we find a least value for SSR

GD' Reduces the Coefficients.

GD finds optimal value for the intercept starting from a R.V.

Our assumed R.V. = 0

Residual = (observed Weight − predicted Weight)^2

$$SSR = (1.4 - (0 + 0.64 \times 0.5))^2$$
$$+ (1.9 - (0 + 0.64 \times 2.3))^2$$
$$+ (3.2 - (0 + 0.64 \times 2.9))^2$$

Gives an eqn of ~~function~~ curve

= 3.1

Derivative of SSR gives the slope of the curve ~~at~~ for a given intercept value.

$$\frac{d}{dx} x^n = n x^{n-1}$$

$$\frac{d}{d\,\text{intercept}}(SSR1) = (1.4 - (\text{intercept} + 0.64 \times 0.5))^2$$

$$= 2(1.4 - (\text{intercept} + 0.64 \times 0.5)) * (-1)$$

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

(-1 = derivative of stuff inside function)

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

$$= (-2)(1.4 - (\text{intercept} + 0.64 \times 0.5))$$
$$+ (-2)(1.9 - (\quad " \quad + \quad " \times 2.3))$$
$$+ (-2)(3.2 - (\quad " \quad + \quad " \times 2.9))$$

GD: Find's min. value by taking steps from an initial guess until it reaches the best value.

Least Square: Finds the ~~min~~ optimal value of intercept where the slope of the curve = 0.

Thus GD is useful when it is not possible to solve for where the derivative = 0

Plug 0 for intercept in the above fn() ⇒

$$\boxed{\frac{d}{d\,\text{intercept}}(SSR) = -5.7}$$

⇒ $\boxed{\text{slope} = -5.7 \text{ when intercept} = 0}$

VIMP Note: The closer we get to the optimal value for the intercept, the closer the slope of the curve gets to 0

⇒ when slope ~~so~~ is close to 0 we need to take

Baby Steps

When slope is far from 0 we take large steps since we are far from optimal value

Step-size is related to slope as it tells regarding 'baby' or 'large' steps to be ~~taken~~

$$\boxed{\text{Step-Size} = \text{slope} * \text{learning rate}}$$

LR = 0.1, 0.01, ...

$\boxed{\text{Used to calculate a new-intercept}}$

New-intercept =
old-intercept - Step Size

∴ new-intercept =

$0 - (-0.57) = 0.57$

$0 \rightarrow 0.57$ (Step-Size change)

(Close to Optimum value)

We Can see how much residual shrinks when intercept = 0.57

now plug in Now intercept
value = 0.57

in the derivative

$$\frac{d}{d\,\text{intercept}}(SSR) = -2(1.4 - (0.57 + 0.64 * 0.5))$$

$$+ \cdots$$

$$= -2.3$$

$\boxed{\text{Now New-slope} = -2.3}$

Step-size $= (-2.3 * 0.1)$
$= -0.23$

New intercept $= 0.8$
$= 0.57 - (-0.23)$

$$\boxed{0.57 \rightarrow 0.8}$$

Compare the residual (gets smaller)

$$\frac{d}{d\,\text{intercept}}(\text{SSR}) = (\text{intercept} = 0.8)$$

$$= -0.9$$

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

$$\text{New intercept} = \text{old-intercept} - \text{stepsize}$$

$$= 0.8 - (-0.09)$$

$$= 0.89$$

Now $\dfrac{d}{d\,\text{intercept}}$ (SSR)

New intercept = 0.89

LR = 0.1

slope	intercept	~~slope~~ $\frac{d}{d\,\text{intercept}}$ (SSR)	step-size	New intercept
0	0.64	~~5.7~~	−0.57	0.57

slope	intercept	slope $\left(\frac{d}{d\,\text{intercept}}(SSR)\right)$	New intercept
0	0.64	−5.7	0 − (−0.57) = 0.57

Step-size = −5.7 × 0.1 = −0.57

New intercept = (0 − (−0.57)) = 0.57

slope = $\cancel{0.57}$ 0.64 ; Intercept = $\cancel{0.64}$ 0.57

$$\frac{d}{d\,\text{intercept}}(SSR) = -2(1.4 + (0.57 + 0.64 \times 0.5))$$

$$+ \ldots$$

Slope = -2.3

Step size = $-2.3 \times 0.1 = -0.23$

New intercept = $0.57 - (-0.23)$

New intercept = 0.80

$$\boxed{0 \rightarrow 0.57 \rightarrow 0.80 \rightarrow 0.89}$$

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

~~$-2(1.4$~~

$-2(1.4 - (0.8 + 0.64 \times 0.5))$

$-2(1.9 - (0.8 + 0.64 \times 2.3))$

$-2(3.2 - (0.8 + 0.64 \times 2.9))$

$\boxed{\text{slope} = -0.9}$

step-size $= -0.9 \times 0.1 =$
-0.09

$$\boxed{\begin{aligned}\text{New-intercept} &= 0.8 - (-0.09)\\ &= 0.89\end{aligned}}$$

Illy new intercept $= 0.92$

" $= 0.94$

" $= 0.95$

$$\boxed{\text{After 6 steps, the GD estimate for Intercept} = 0.95}$$

The LSE for intercept is also 0.95

GD stops

→ when step-size ≈ 0
i.e, when slope is ≈ 0

→ In practice the min step-size $= 0.01$ or smaller some

→ no. of steps = 1000 (or >)

→

---

(1) Calculate SSR

(2) $\frac{d}{d\,\text{intercept}}$ (SSR) to find slope
= derivative of loss func,

(3) pick a R.V. for intercept
(=0 in our case)

(4) Calculate derivative when intercept = 0

(5) Result value is plugged for
step-size calculation
= slope × LR

(6) New intercept = old-intercept − step-size

(7) Now plug in new intercept
into $\frac{d}{d\,\text{intercept}}$ (SSR) & ---

until step size is close to 0.

How to estimate intercept & slope

we will use SSR as loss func

<u>3-axis:</u>

Axis 1 = slope

" 2 = intercept

" 3 = SSR

we want to find values for slope & intercept that gives min SSR

Take derivative of loss for c)

(1) Take derivative w.r.t. intercept
(2) " " slope

$$\frac{d}{d\,intercept}(SSR) = \frac{d}{d\,intercept}\left(1.4 - (intercept + slope * 0.5)\right)^2$$

$$\left(slope = const\right) \quad (= -1)$$

$$\therefore \frac{d}{} = -2\left(1.4 - (intercept + slope * 0.5)\right)$$

$$+ (-2)\left(1.9 - (intercept + slope * 2.3)\right)$$

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

$$\frac{d}{d\,slope}(SSR) = 2\left(1.4 - (intercept + slope * 0.5)\right) \times 0.5$$

$$+ 2(\;---\;) \times 2.3 + 2(---)\,2.9$$

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

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

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

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

Gradient:

When u have 2 or more derivatives of same func, they are called "**Gradient**".

We use Gradient to descend to the lowest point in the loss func, which is SSR.

∴ This algo is called **GD**.

(1) Pick a Random # for intercept = 0

(2) Pick a random no. for slope = 1

Starting points:

$\text{Slope} = \frac{y}{\cancel{0}}$; Intercept = 0.

Now plug the values into $\frac{d}{d\,\text{intercept}}(SSR)$ &

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

c) $\frac{d}{d\,\text{inter}}(SSR) = -1.6$

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

$\text{step-size}_{\text{intercept}} = \text{slope} * LR$

$= -1.6 * 0.01$

$= -0.016$

" (slope) $= -0.8 * 0.01$

$= -0.008$

GD is highly sensitive to LR

LR = start with a large value & gradually decrease

Now new-intercept

$= \text{old-intercept} - (\text{step-size})$

$= 0 - (-0.0016)$

$= 0.0016$

New slope = old slope − stepsize

$= 1 - (-0.008)$

$= 1.008$

Repeat this process until all step-sizes are very small or reach the maximum of steps

Best fitting - line

Intercept = 0·95

Slope = 0·64

Same values are got from least squares.

(1) Take derivative of each loss func for each parameter in it.

i.e. take Gradient of loss func

(2) Pick RV's for parameters

(3) plug parameter values into derivatives (Gradient)

(4) Calculate Step Size

(5) Calculate new parameters

New parameter = old parameter − Step size

Now go back to step 3
& repeat until step size
is very small or
U reach max. no of steps.