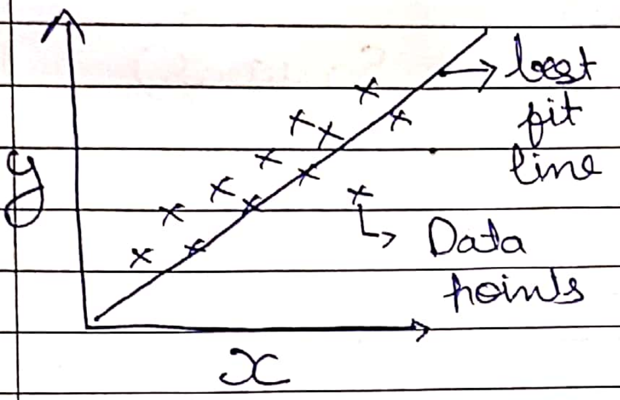
Linear Regression



Simplest Linear regression is of the form

$$\boxed{\underset{\text{target}}{y} = \underset{\text{slope or weight}}{m}\,\underset{\text{feature}}{x} + \underset{\text{intercept/bias}}{b}}$$

This was for a dataset having one feature, one target. Such datasets are rarely used in practice (Time series is similar but different).

So let expand our intuition when we have multiple features.

$$\underset{\text{target}}{y} = \theta_1 x_1 + \theta_2 x_2 \;\cdots\; \theta_n x_n = \underbrace{X^T\theta}_{\text{vectorized implementation}}$$

slope/ weights — θ_1, θ_2

features — $x_2 \cdots x_n$

$x_1 = \cancel{\theta}\ 1$

↳ bias term

Scanned with CamScanner

Two ways to update weights and biases are:

1) OLS (Oridinary Least Squares)
2) Gradient Descent.

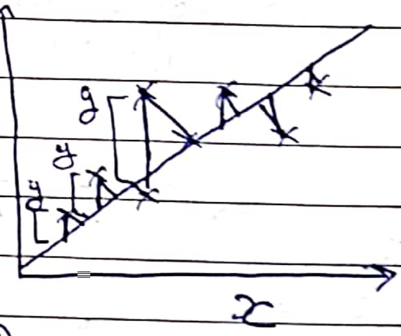
Ordinary Least Squares

Slope $\boxed{m = \dfrac{\sum_{i=0}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=0}^{n}(x_i - \bar{x})^2}}$

bias $\boxed{b = \bar{Y} - m\bar{X}}$ where $\bar{x}, \bar{y}$ = mean

Lets now prove these:

$$\text{error} = (y_1 - \hat{y})^2 + (y_2 - \hat{y})^2 \cdots + (y_n - \hat{y})^2$$



$$\text{error} = \sum_{i=1}^{n}(y_i - \hat{y})^2 \quad \text{—} (1)$$

But we know $\hat{y}_i = mx_i + b_i$ —(2)

Using (2) in (1)

$$\text{error} = \sum_{i=1}^{n}(y_i - (mx_i + b_i))^2$$

To reduce this error, we have to differentiate the equation wrt to m and b (as they can only reduce the error)

Chitra

"It is the prime responsibility of every citizen to feel that his country." —Sardar Vallabhai Patel

Scanned with CamScanner

So

$$\frac{d}{db}\sum(y_i-(mx_i+b_i))^2=0$$

and

$$\frac{d}{dm}\sum(y_i-(mx_i+b_i))^2=0$$

Lets solve this

$$\sum 2(y_i-(mx_i+b_i))(-1)=0$$

~~we~~

$$\sum(y_i-(mx_i+b_i))=0$$

this is just nuber of features × bias

$\sum b_i = nb$ — (3)

using (3) we get

$$\sum y_i-(mx_i+nb)=0$$

$$\sum y_i-\sum mx_i - nb=0$$

Divide by n we get

$$\frac{\sum y_i}{n} - m\frac{\sum x_i}{n} - b = 0$$

$$\bar{y} - m\bar{x} - b = 0$$

$$\boxed{b=\bar{y}-m\bar{x}}$$

Scanned with CamScanner

Date ___/___/_______ Page No.: _______

Now,

$$\frac{d}{dm} \Sigma (y_i - (mx_i + b_i))^2 \quad -①$$

Now from previous we know

$$b_i = \bar{y} - m\bar{x} \quad -②$$

Lets use ② in ①

$$\frac{d}{dm} \Sigma (y_i - (mx_i + \bar{y} - m\bar{x}))^2 = 0$$

$$\Sigma 2(y_i - (mx_i + \bar{y} - m\bar{x}) \cdot (x_i - \bar{x})) = 0$$

$$\downarrow$$

$$\Sigma (y_i - (mx_i + \bar{y} - m\bar{x}) (x_i - \bar{x}) = 0$$

$$\Sigma (y_i - \bar{y}) - m(x_i - \bar{x})) (x_i - \bar{x}) = 0$$

$$\downarrow$$

$$\Sigma (y_i - \bar{y})(x_i - \bar{x}) = m \Sigma (x_i - \bar{x})^2$$

$$\downarrow$$

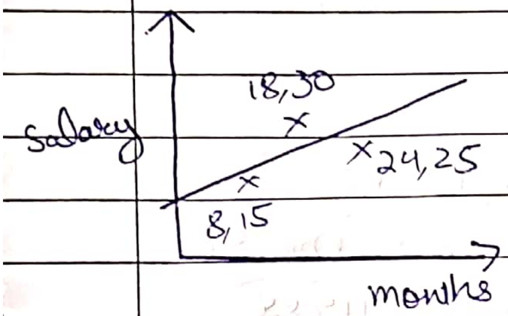
$$\boxed{m = \frac{\sum_{i=0}^{n} (y_i - \bar{y}) (x_i - \bar{x})}{\sum_{i=0}^{n} (x_i - \bar{x})^2}}$$

Scanned with CamScanner

Date ___/___/___ Page No..

Gradient Descent

- It is an optimization technique

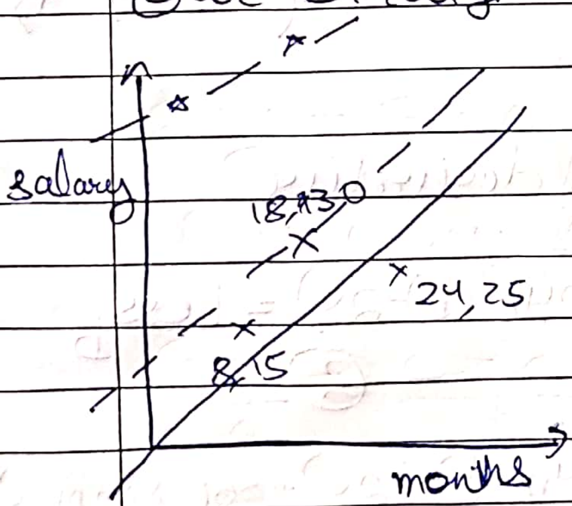


we have to find the line ($y=mx+b$) which reduces the residual error sum.

For now lets assume m is fixed ; $\boxed{m=1}$

$$error = (y_1-\hat{y}_1)^2 + (y_2-\hat{y}_2)^2 + (y_3-\hat{y}_3)^2 \quad -(1)$$

Our straight line equation at $m=1$ is $y=x+b$



-*- $m=1, b=0$

--- $m=1, b=5$

—— $m=1, b=20$

b	error
0	194
5	69
⋮	
20	500

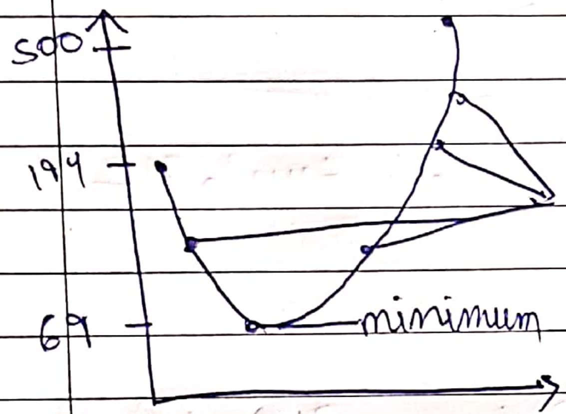
we calculated error using (1) since we know both m and b and actual $\hat{y}$

"Education is the manifestation of perfection already in man." —Swami Vivekananda

Chitra

Scanned with CamScanner

If we plot this bias and error (Loss) we get



values we get if we solve for more we get this curve

Lets do some math!

$$Loss = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

So for our salary graph:

$$Loss = ((8m+b)-15)^2 + (24m+b)-25)^2 + ((18m+b)-30)^2$$

we start filling some values

Lets calculate the partial derivative

$$\frac{df(m,b)}{db} = 2((8m+b)-15) \times (1) + 2((24m+b)-25) + 2((18m+b)-30) = Loss_b \quad -(2)$$

$$\frac{df(m,b)}{dm} = 2((8m+b)-15)\cdot 8 + 48((24m+b)-25) + 36((18m+b)-30) \quad -(3)$$

$$= Loss_m$$

Scanned with CamScanner

Date ___/___/___ Page No.. ___

Initialise m, and b with some value say

$m=1$, $b=2$

and suppose we get $Loss_b = -20$

$Loss_m = -13$

I multiply $Loss_b$ and Loss m with α (Learning rate), Let α be 0.01

$Step_b = -20 \times 0.01 = -0.02$

$Step_m = -13 \times 0.01 = -0.013$

we get new value for b and m

$n_b = old_b - Step_b = 2 - (-0.02) = 2.02$

$n_m = old_m - Step_m = 1 - (-0.013) = 1.013$

we keep updating this to reach the min