**Linear Regression**

Case1:

Saipraasad hired as a Manager for Nykaa

At that time of he joins Nykaa sales very low, his main goal to increase the sales in next 3 months.

He asks the Zubair ===== marketing

1. NP
2. TV
3. SM

Sales= 2\*NP+5\*TV+10\*SM ========== Nithin as a DS

He ask the qn to Nithin , is my equation correct?

| NP | TV | SM | Sales |
| --- | --- | --- | --- |
| 200 | 300 | 500 | 2000 |
|  |  |  |  |

One Project

PM TL Dev

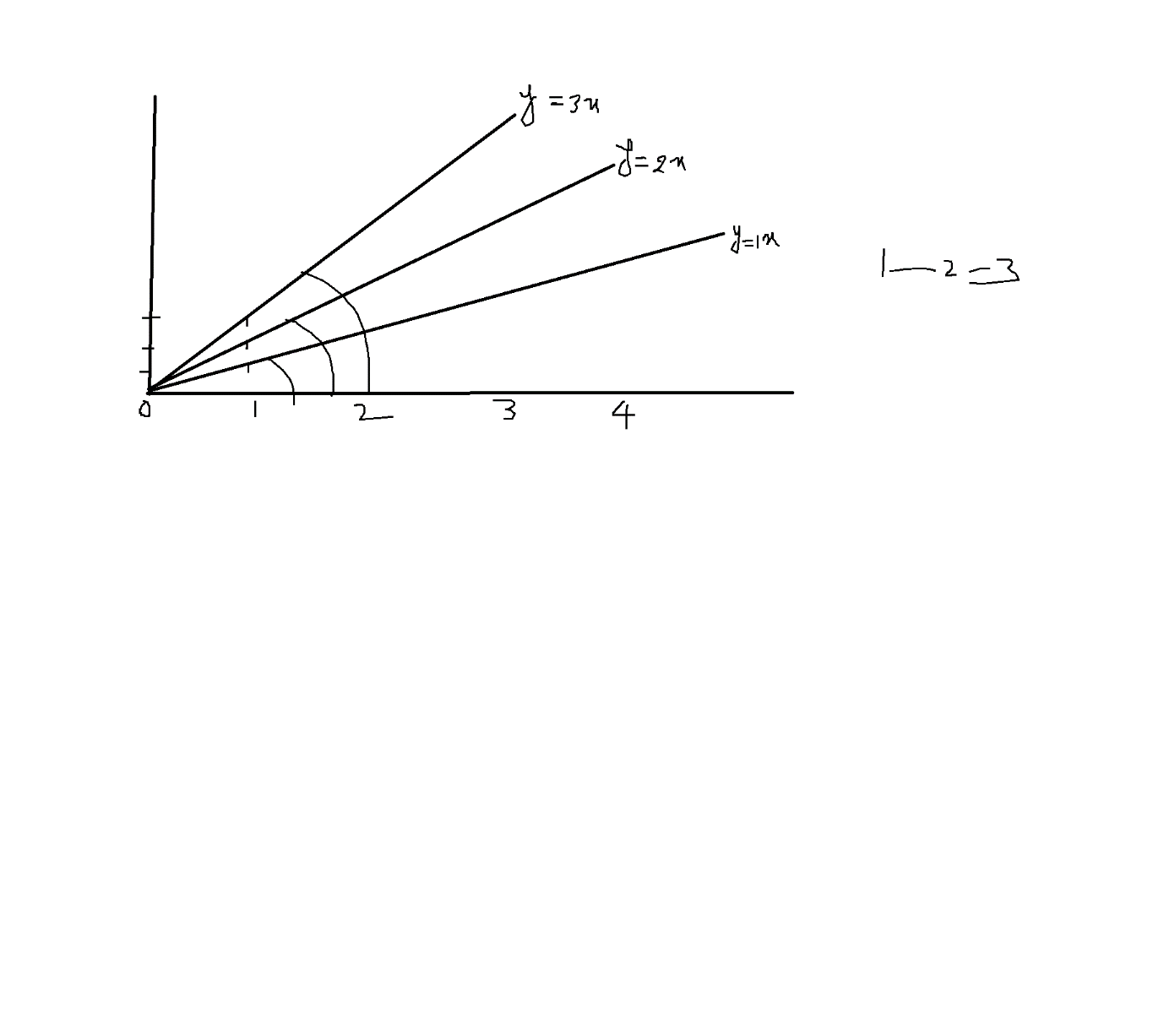
Success of project=1\* Pm+ 1\* Tl+1\* Devloper

Sales= k1\*NP ======= > y=mx

What Is m ?

slope

| X | Y=x | Y=2x | Y=3x | Y=3x+2 |
| --- | --- | --- | --- | --- |
| 0 | 0 | 0 | 0 | 2 |
| 1 | 1 | 2 | 3 |  |
| 2 | 2 | 4 | 6 |  |
| 3 | 3 | 6 | 9 |  |
| 4 | 4 | 8 | 12 |  |



How much the line should be rotated will provide by slope value m

The slope value will change ====== rotation of line will also change

How much you want lift the line will provide by intercept (c)

Y=mx+c

Normal maths

Machine learning

Deep Learning/AI:

His money on TV NP SM======= Inputs

Sales ============ Outputs

We have 3 input variables

X1 x2 and x3

Y= b1\*x1+b2\*x2+b3\*x3+b0

B0= bias

B1 b2 b3 are coefficients

X1 x2 x3 are input

Y output

Normal maths

Machine learning

Deep Learning/AI:

Linear =========== Non linear

Linear =========== Power 1

Non linear ========= Power >1 n

N=2 ======== > parabola

Ax+by+c=0

By=-AX-C

Y= -(A/B)\*X- C/B

Y= Mx +C

Why we are interesting the line equation

Coeff

Rotation of the line

Data ======== > 100 observations

| NP (rs) | TV (rs) | SM (rs) | Sales |
| --- | --- | --- | --- |
| 200 | 300 | 400 | 2000 |
| 150 | 250 | 500 | 1500 |
| 300 | 400 | 500 | 3000 |
| 120 | 220 | 320 | 1200 |
| 150 | 300 | 450 | 2500 |

Step-1: we divide our data into two parts

Train

Test

I will divide 100 observation 70:30 80:20 60:40

70:30

Out of 100 observations 70 observations are train data

Out of 100 observations 30 observations are test data

Note: Train data should be high

10k ========= > 8k +2k 80:20

10lakhs ==== 9.9laks train + 10k test 99:1

1cr ======= 99 laks train + 1lak

1000 : 70:30

700 +300

800+200

900+100

Train

NP TV SM sales

1

2

3

.

.

.

70

Test

NP TV SM Sales

71

72

73

.

.

.

100

We divide data into randomly

Step-2:

Train data has both input and output

Machine learning model will develop by using train data

Step-3: Model is ready

Then will pass test input data

But test data has output data also ===== actual values

===== actual sales

(NP ,SM,TV) ======= Model ========= predicted values

========= predicted sales

Step-4:

| NP | SM | TV | sales | Model=  Y=b0+b1\*NP+b2\*SM+b3\*TV  Predicted values |
| --- | --- | --- | --- | --- |
| 500 | 400 | 200 | 3000 | 3500 |

500,400,200 ========= Model ======= 4000

The model output predicted values will compare with test output data(actual values)

Y=100+10\*500-4\*400+2\*200

Y= 100+5000-1600+400= 5500-1600= 3900

Train Test Input Output Actual predicted values

**Goal In Linear regression:**

**Identify the best line**

**By identify the best coefficients**

**Finding the suitable coefficients in order to reduce the error**

1. **Error:**

**Actual value- predicted value**

| NP | SM | TV |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 500 | 400 | 200 | 3000 | 3200 | 3000-3200=-200 |
| 400 | 550 | 250 | 2800 | 2600 | 2800-2600=200 |

1)range

2) Mean Deviation

3) Absolute mean

4) Variance

5) Standard deviation

