In order to understand the data

1. Central tendency

* Mean
* Median
* Mode
* Difference between mean and median
* Skewed distributions
* Pos/neg/no

All maths developed by assuming data follows normal distribution

Mean= median =mode

Generally data never follows normal distribution

Transformation techniques

We need to learn outlier analysis

Data skew will happen bcz of outliers

1. Data dispersion

* Range
* Mean deviation
* Absolute mean deviation
* Variance
* Standard deviation

1. Range:

1,2,3,4,5,6,…..,100

Which point to which point, the data is distributed

Lower value=1

Upper value =100

Range= U.V-L.v= 100-1= 99

1 ------- 2 ----------- 3 --------- 4

4-1= 3

Drawback:

* Range never consider middle value
* It will consider only lower and upper values

1. Mean deviation:

How much Each every data point is deviated from mean point

Calculate the mean

* Mean deviations from x1 :

X1 is deviated by 2 units from mean of the data

* Mean deviations from x2 :

X1 is deviated by 2 units from mean of the data

* Mean deviations from x3 :

X1 is deviated by 2 units from mean of the data

* Mean deviations from x4 :

X1 is deviated by 2 units from mean of the data

* Mean deviations from x5 :

X1 is deviated by 2 units from mean of the data

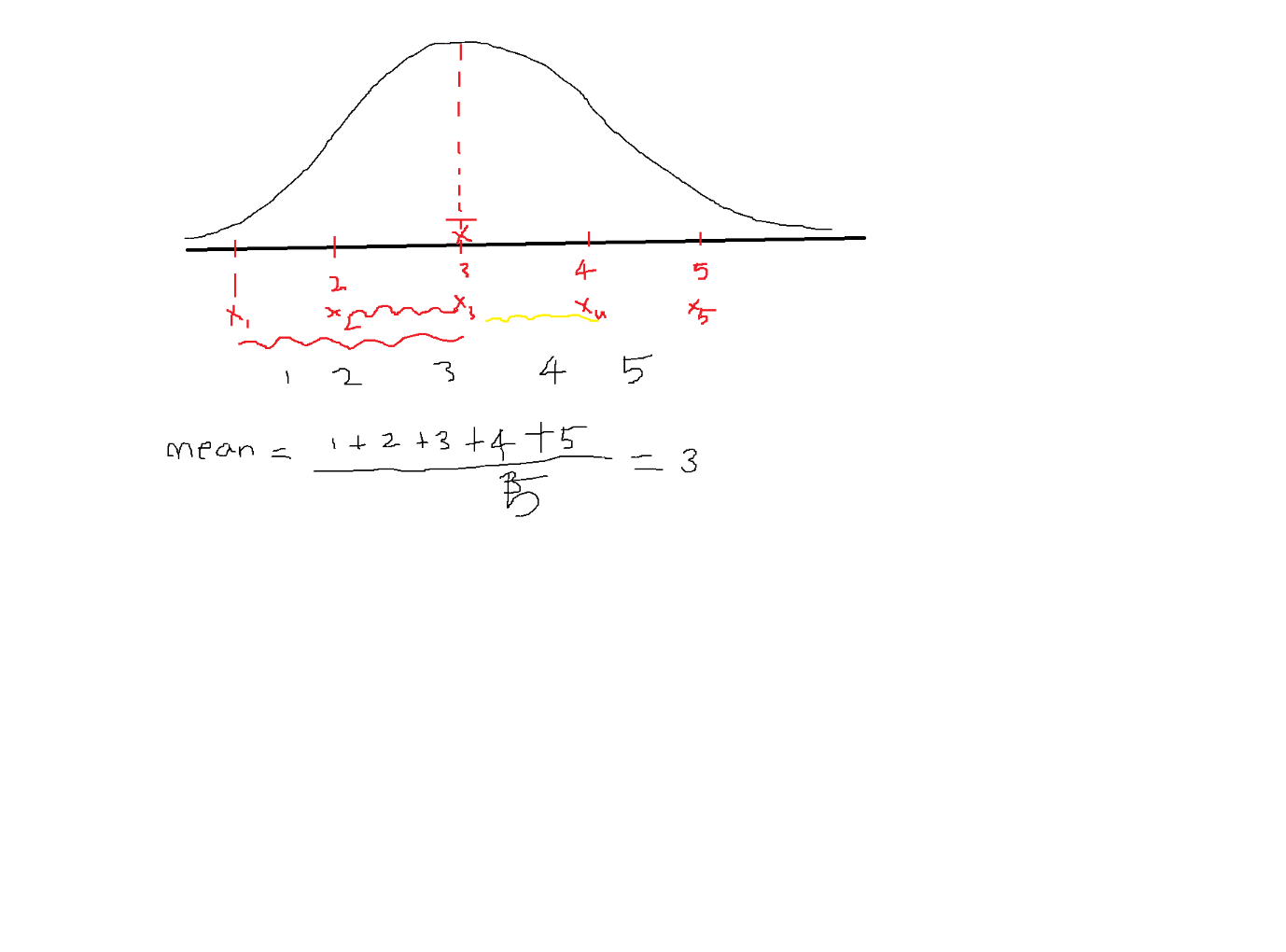
Total deviation of 5 observations

(

Total deviation of n observations

(

(



Draw back:

For the total 5 observations

The total mean deviation=

(

We are seeing the deviation w.r.t each point

But the total deviation says : There is no deviation

Hence math fail

Yes =1 No=0----------- > 1-0=1

No =0 Yes=1 -------- > 0-1=-1

Yes =1 No=0 -------- > 1-0=1

No =0 Yes=1 ------- > 0-1 =-1

Total error = 1-1+1-1=0

This is a great ====== it is working and understanding each and every data point

Why it is happens zero ===== > positive and negative cancellations

+2 ====== > 2 points ahead from mean

-2 ====== > 2 points below from mean

I=-5Amp

How to avoid + and –

Which concept in maths will convert – to +

1. Mod /absolute values
2. Multiply \*
3. Square
4. Absolute mean deviation:

* Absolute Mean deviations from x1 :

X1 is deviated by 2 units from mean of the data

* Absolute Mean deviations from x2 :

X2 is deviated by 2 units from mean of the data

* Absolute Mean deviations from x3 :

X3 is deviated by 2 units from mean of the data

* Absolute Mean deviations from x4 : |

X4 is deviated by 2 units from mean of the data

* Absolute Mean deviations from x5 :

X5 is deviated by 2 units from mean of the data

Total deviation of 5 observations

|

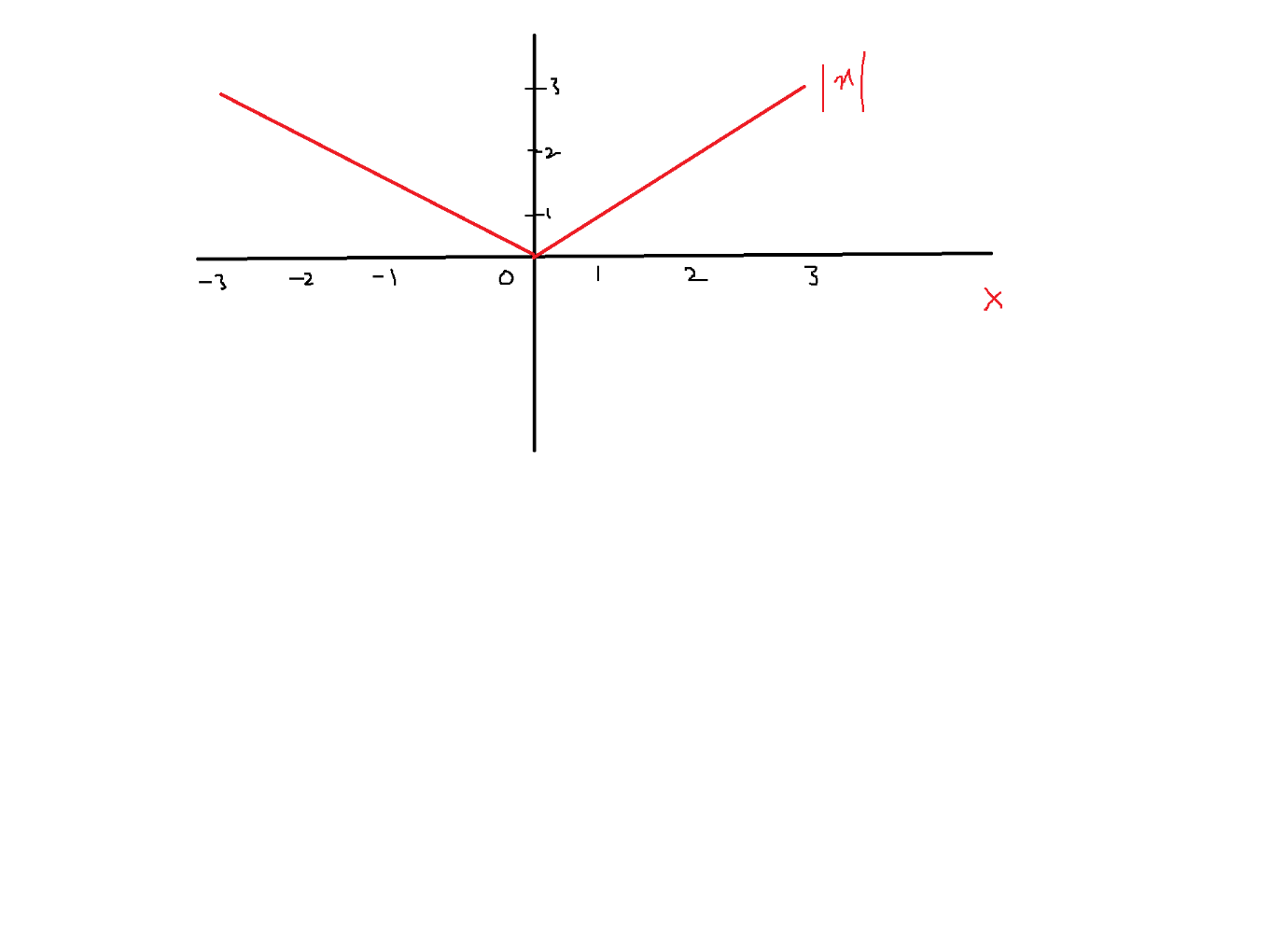
Total deviation of n observations

|

|

Draw back:

| x | Y=|x| |
| --- | --- |
| -3 | 3 |
| -2 | 2 |
| -1 | 1 |
| 0 | 0 |
| 1 | 1 |
| 2 | 2 |
| 3 | 3 |

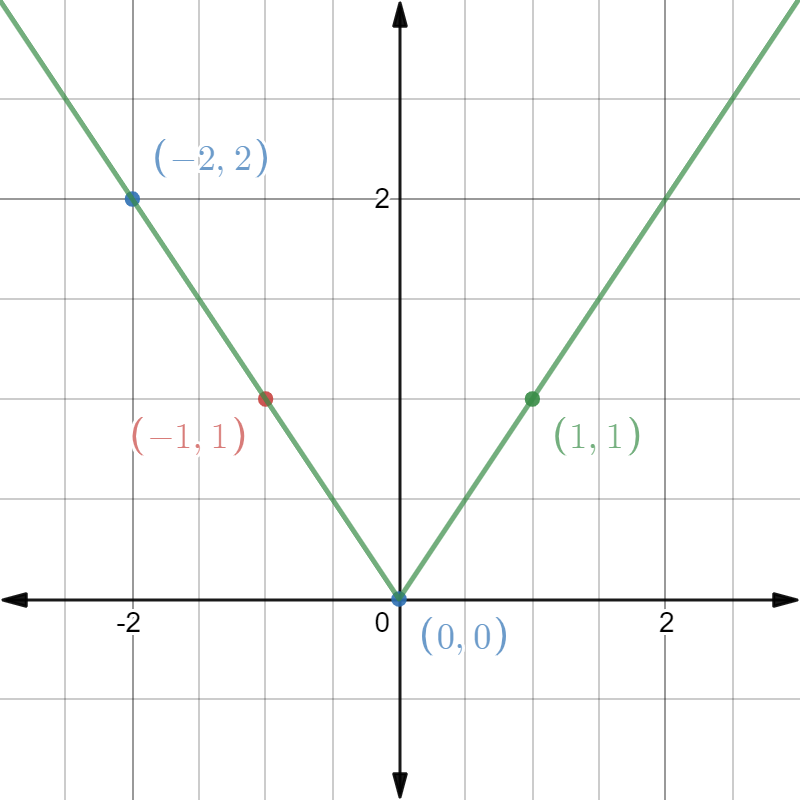


|x| graph is discontinue at point zero

Whenever any equation discontinue at some point

Then

Differentiation is not possible



What is |x| differentiation=

d(|x|)/dx = x/|x|,

at x=0 ======= > 0/0 ========= >

indeterminant form

math fail

* Mean deviation gives total deviation as zero
* To avoid that we taken absolute value
* Absolute value giving total deviation
* The graph of |x| is a V-shape
* It is discontinue at point zero
* Anything discontinue at zero , some math concepts fail
* Differentiations and integration will not work
* |x| differentiation = x/|x|
* Keep x=0 ------ > 0/0 ----- indeterminant form
* To +

One option ======== > absolute

Another option = square

-1\*-1= 1

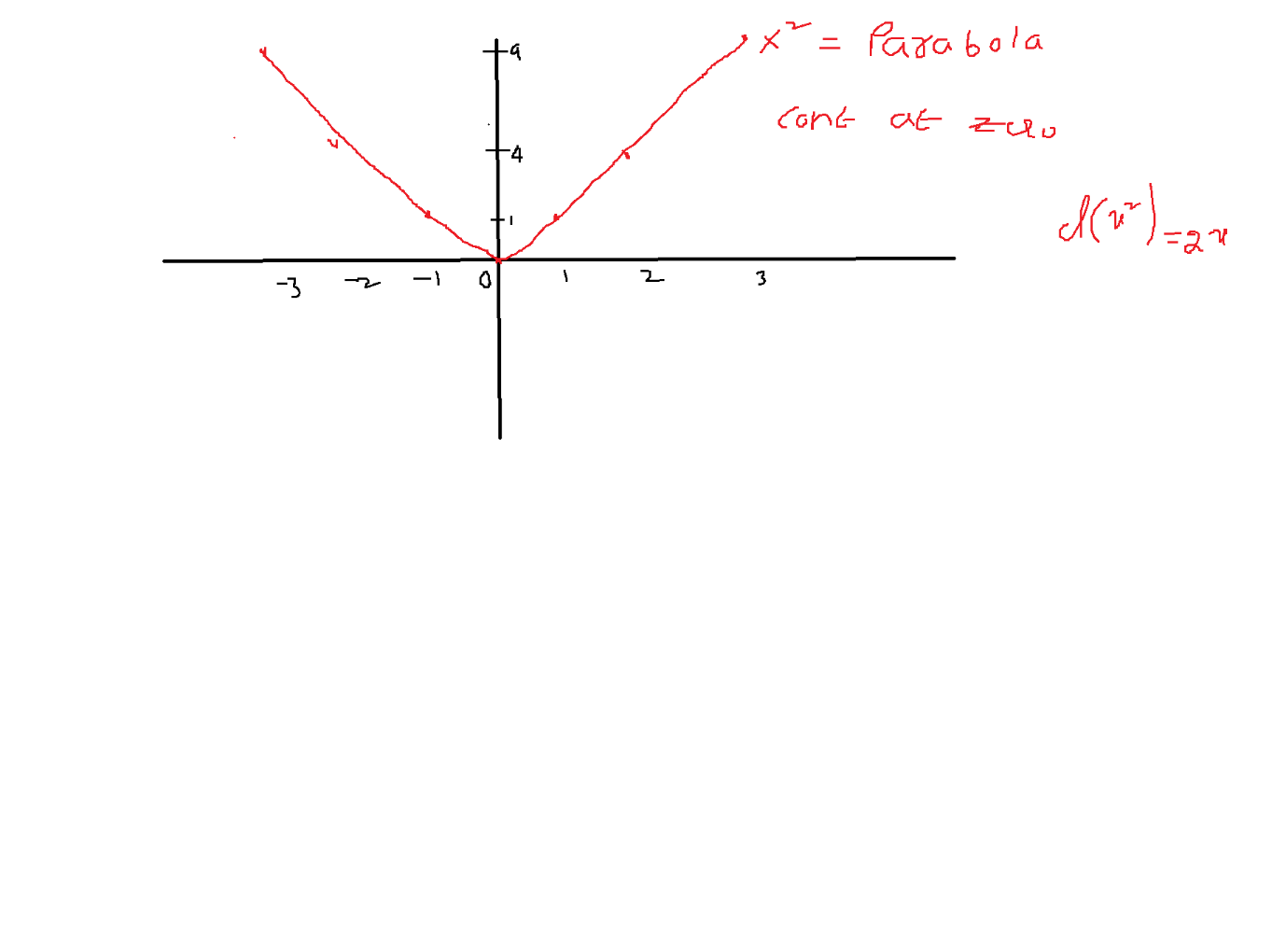
1. Variance:

Total deviation of n observations

(

Draw back:

| x | Y=x^2 |
| --- | --- |
| -3 | 9 |
| -2 | 4 |
| -1 | 1 |
| 0 | 1 |
| 2 | 4 |
| 3 | 9 |
|  |  |



* The values will be rise , because of squaring

Km fare

1km 5rs ========== > (1km-3km)^2= -2km^2=4km^2

2km 10rs

3km 20rs

4km 30rs

5km 40rs

1+2+3+4+5/5=3km

Not only values units also rise

Km^2 =========== km?

Square root (km^2) ====== km

1. Standard deviation

1. Range: H-L