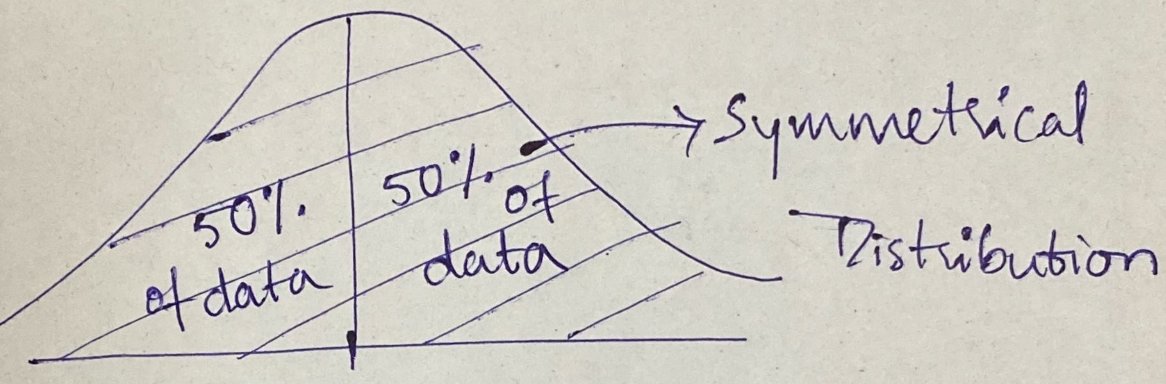
DAY - 3

1.) Normal Distribution (or) Gaussian :

(Bell-curve)



Age, weight, height follows Gaussiandistibution by using kernal density Estimator (kde) we can smoothen the histogram and create the curve.

* Emperical Rule of Normal Distribution :

Emperical Formula - $\boxed{68 - 95 - 99.7\%}$

Within 1^{st} standard Deviation from left & right 68% of Data will be present.

2^{nd} σ from left & right 95% of Data.

3^{rd} σ from left & right 99.7% of Data

We can make perfect Assumptions from Guassian Distribution.

2) Standard Normal Distribution :

Let 'x' be a variable (mean)(std)

x = Gaussian distribution (μ, σ)

We can transform x to y ↓ How ?

Where, y = SND $[\mu = 0, \sigma = 1]$

$$Z\text{-score} = \frac{x_i - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$\frac{\sigma}{\sqrt{n}}$ — standard error

n will be = 1, because we gonna apply it everywhere (not always n will be 1)

when $n = 1$ the Z-score value will be as

$$Z\text{-score} = \boxed{\frac{x_i - \mu}{\sigma}}$$

But the Question is why we are converting x with variable y?

$x = \{1, 2, 3, 4, 5\}$

$\mu = 3$, $\sigma = 1.414$

$y = \left\{ \frac{(1-3)}{1.414} = -1.414, \frac{(2-3)}{1.414} = -0.707, 0, \frac{(4-3)}{1.414} = 1.414, \frac{(5-3)}{1.414} = 0.707 \right\}$ → by applying Z-Score formula

So, no we can see the drastic change in variable y-values

Why?

(years) Age	(Kg's) Weight	(cm's) height
24	72	150
26	78	160
32	84	165
33	92	170
34	87	150
28	83	180
29	80	175

[At the end of the day we apply m.l algorithms to machine. those algorithm is nothing but Maths formula]

By above table we can assume that our math-calculation time will be higher because of unit scale is different and values also differs by huge number.

if we apply Z-score value to Age, height and weight all the values will be on same scale and it will be from -3 to 3 with $\mu = 0$
Calculation will also be take less time
This Entire process is called as "Standarization".

Also, we can revert back the scale after training the model with same Z-score formula

Stdarization of Age, Height, Weight is
$\mu = 0$, $\sigma = 1$ ☆

* Normalization: We try to normalize the value between lower scale to higher scale (Transformation to 0 to 1)

(i) Min-Max Scaler
↓

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

x	y
1	0
2	0.25
3	5
4	0.75
5	1

$\frac{1-1}{5-1} = 0$, $\frac{2-1}{5-1} = \frac{1}{4}$, $\frac{5-1}{5-1} \rightarrow 1$

$\frac{3-1}{5-1} = \frac{2}{4}$, $\frac{4-1}{5-1} = \frac{3}{4}$

$\boxed{0 \text{ to } 1}$

We can apply Normalization in D.L & M.L

Ex: We have a b/w picture with pixel 0-255, we can change into 0-1 values, this is called normalization.

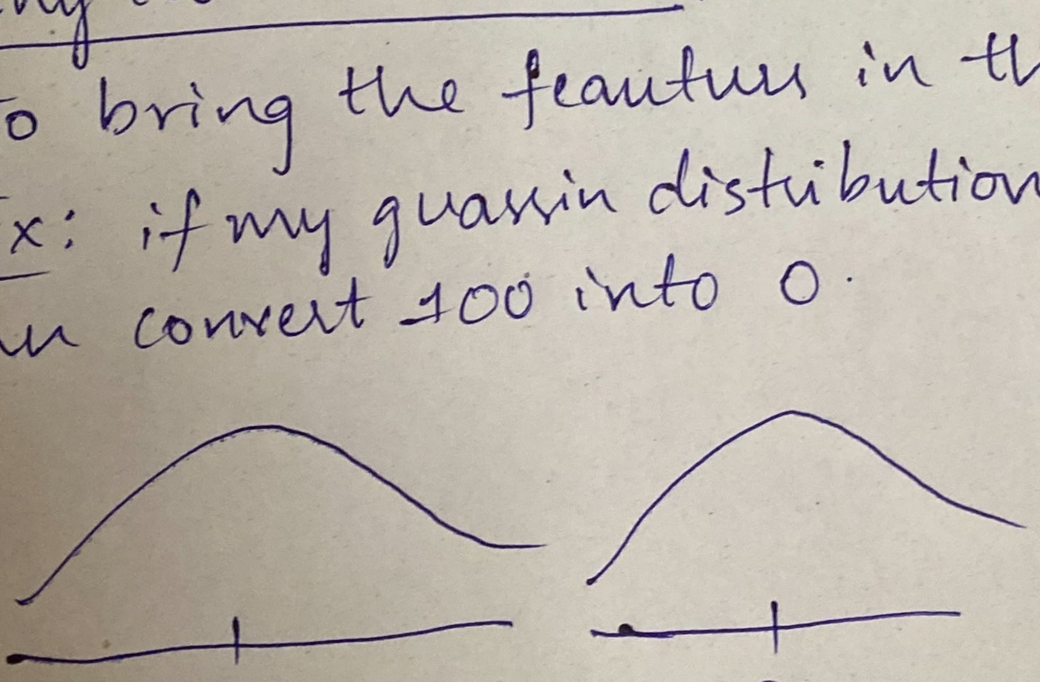
Standardization:

We have a variable 'x' with values (μ & σ) which is normal distribution now, i'll apply Z-score formula to 'x' and i'll get another values in another variable 'y' which is standard normal distribution where, my $\mu = 0$ & $\sigma = 1$

Why do we do this?

To bring the features in the same scale

Ex: if my guassin distibution is 100, I can convert 100 into 0.



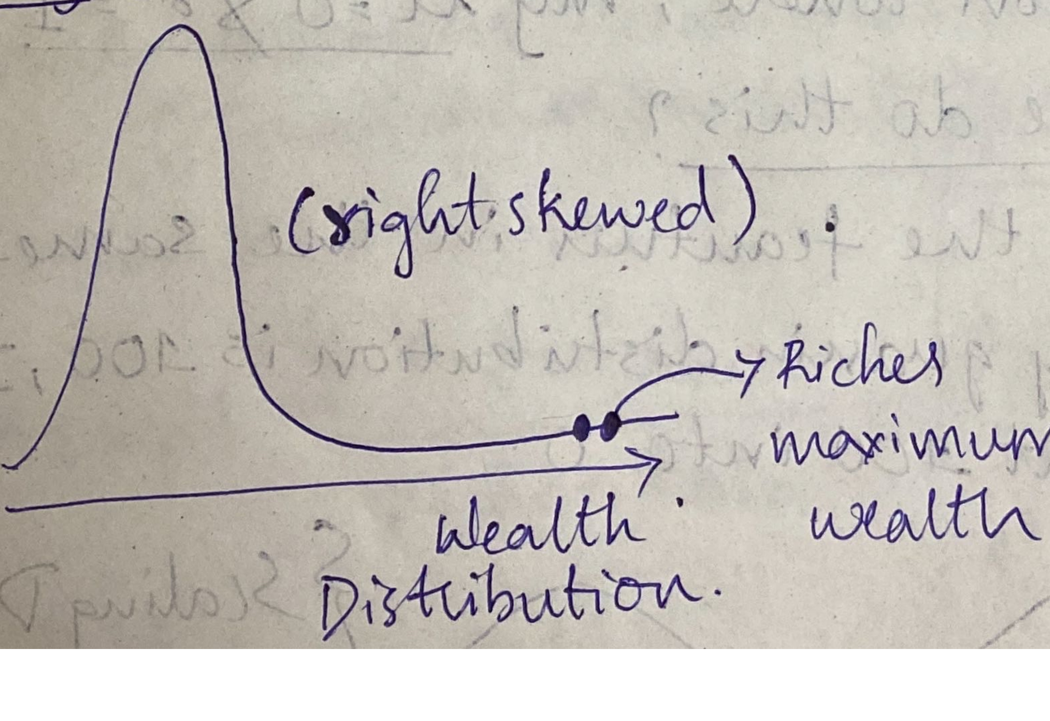
Normalization: Here, we are not giving $\mu = 0$, $\sigma = 1$ but we ourselves are giving some range maybe [0-1] using **min-max scaler**. We are normalizing the data between given fixed data/values.

We can use standardization & Normalization to scale down the values which will be used in ml & dl.

In, Rnn, Ann we use standadization

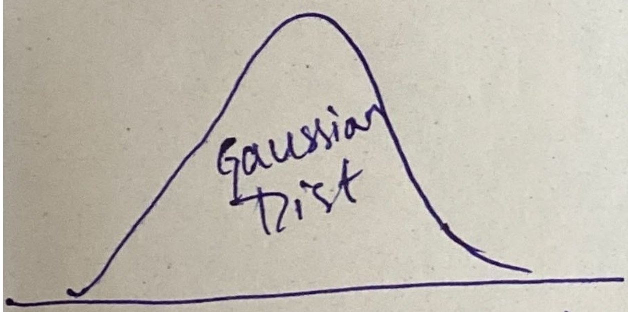
In cnn we use Normalization.

5) **Log Normal Distribution**:

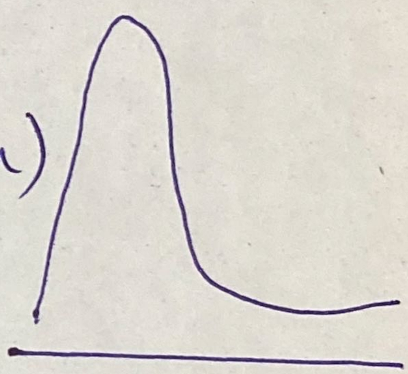


If the random variable 'x' is log - normally distributed, then y = ln (x) (natural log) has a normal Distribution.

So, Let's consider 'x' is log normal Distribu
I will apply natural log of x on another variable 'y', in y if we get Guassian Distribution we can prove x is log normal distribution.



(Antilog)
$\Longrightarrow$ exponentia(x)
$\hookrightarrow$ (y)



$x = nd(\mu, \sigma)$

(only applied to rightskewed)

Natural log = $\log_e (y = \ln(x))$

We can use z - table to know the exact value of area under curve.

There are two types :

Negative z - Score table
positive z - Score table