Subject: Data Science

MON TUE WED THR FRI SAT SUN

10th September, 2022

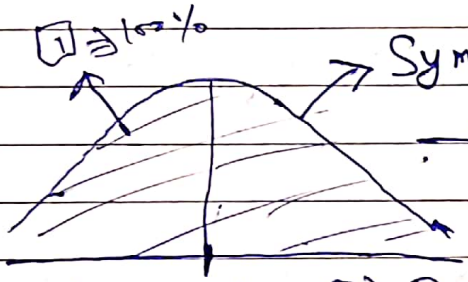
Statistics Note (Day-3)

By Krish Naik Sir

Agenda

1. Normal Distribution
2. Standard Normal Distribution.
3. Z-Score
4. Standardization And Normalization.

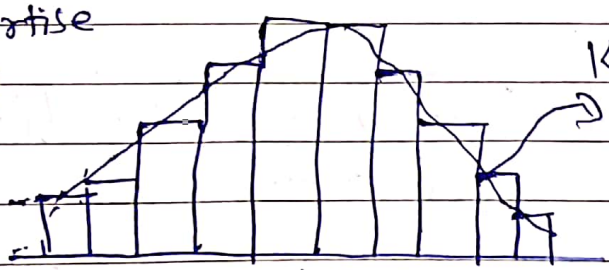
1. Normal Distribution / Gaussian:-



Eg: Age, weight, height → Distributions

Domain Expertise (Doctors)

Eg:



Eg:

IRIS Dataset

⇓

(Petal length, Sepal length, petal width, Sepal width)

⇓

Gaussian Distribution.
(will be follow)

your writing partner

Always Refer wikipedia Search :-

Subject

MON TUE WED THR FRI SAT SUN

Emperical Rule of Normal Distribution

Gaussian / Normal Distribution

↓

Assumptions of the data //

Experical Formula / Rules :-

→ 68 - 95 - 99.7 % Rule //

Q-Q Plot ⇒ Distribution is Gaussian

Not?

Standard Normal Distribution :-

(Note :- μ = means, σ ⇒ Standard Deviation)

$X \approx$ Gaussian Distribution (μ, σ)

↓

$y \approx SND \ (\mu = 0, \sigma = 1)$

Converting X Distribution to y Distribution by using

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

your writing partner

Subject _____

MON TUE WED THR FRI SAT SUN

Eg $X = \{1, 2, 3, 4, 5\}$

$\mu = 3$

$\sigma = 1.41$

let,
μ = mean
σ = Standard Deviation
n = Sample size

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

$\Rightarrow$ Standard Error

$\Downarrow$

Inferential Stats.

let, $n = 1$

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

Eg :- $X = \{1, 2, 3, 4, 5\}$

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

$$= \frac{1-3}{1.414} = -1.414$$

$$= \frac{2-3}{1.414} = -0.707$$

$$= \frac{4-3}{1.414} = \frac{1}{1.414} = 1.414$$

your writing partner

$\Rightarrow$ $y = \{-1.414, -0.707, 0, 0.707, 1.414\}$

Subject _____

MON TUE WED THR FRI SAT SUN

$y = \{-1.414, -0.707, 0, \ldots\}$

Why ?? [Standardization] $\Rightarrow \mu = 0, \sigma = 1$

Eg

	Age (years)	Weight (Kg)	Height (cm)
$\mu = 0$	24	72	150
$\sigma = 1$	26	78	160
	32	84	165
	33	92	170
	34	87	150
	28	83	180
	29	80	175

Same scale

Machine Learning

Math Equations

Algorithm $\Rightarrow$ Mathematical Model

∴ (mathematical calculation take more Time)

Subject _____

MON TUE WED THR FRI SAT SUN

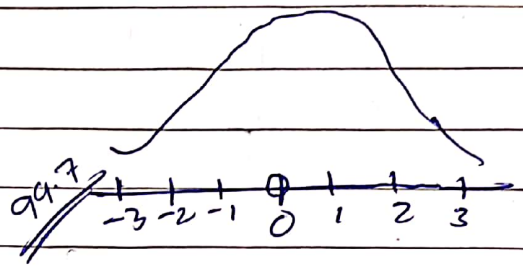
Feature Scaling or Normalization

Standarization [z-score]

$[0-1]$ $[-1, 1]$

$[\mu = 0, \sigma = 1]$

$[-3 \leftrightarrow 3]$



Normalization $\Rightarrow$ [lower Scale $\leftrightarrow$ Higher Scale]

Min Max Scaler [0 -1]

$$x_{scaled} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Subject _____

MON TUE WED THR FRI SAT SUN

Min Max Scaler [0-1]

$$X_{scaled} \Rightarrow \frac{X - X_{min}}{X_{max} - X_{min}}$$

$\frac{1-1}{5-1} = 0 \Rightarrow 0$

$\frac{2-1}{5-1} = \frac{1}{4} \Rightarrow 0.25$

$\frac{3-1}{5-1} = \frac{2}{4} \Rightarrow 0.5$

$\frac{4-1}{5-1} = \frac{3}{4} \Rightarrow 0.75$

$\frac{5-1}{5-1} = \frac{4}{4} \Rightarrow 1$

Eg

x	y
1	0
2	0.25
3	0.5
4	0.75
5	1

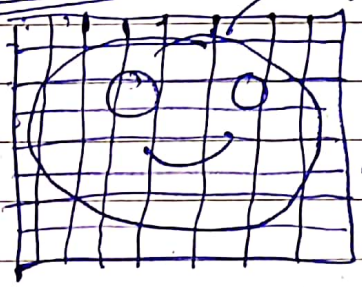
0 to 1

0 to 1

Apply ??

Deep Learning

Where Images Apply (0-255 pixels)



0-255 pixel convert to 0-1 pixel

Normalization

Scale Down

Eg BE PHD NASA

B.E B.E B.E

For communication we have to

Subject _____

MON TUE WED THR FRI SAT SUN

Revision

① Standardization :-

$$Z\text{-Score} = \frac{X_i - \mu}{\sigma}$$

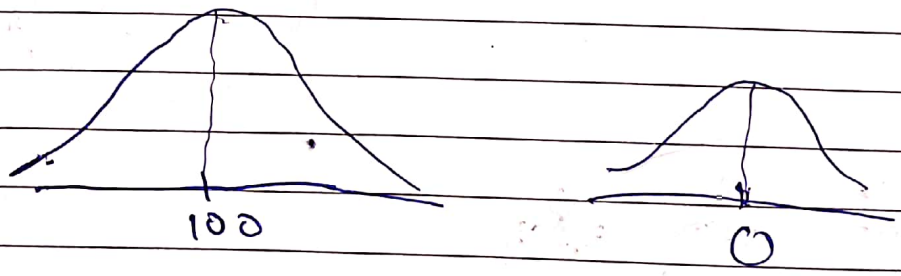
$X \longrightarrow$ Normal Distribution (μ, σ)

$\Downarrow$ Z-score

$y \Rightarrow$ SND $(\mu = 0, \sigma = 1)$

⭐ Why do we do this :-

→ Bring the feature in the same Scale.



Scaling Down

your writing partner

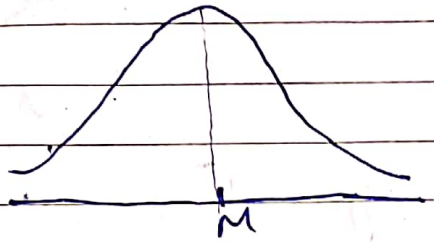
Subject _____

MON TUE WED THR FRI SAT SUN

(2) Normalization [0-1]

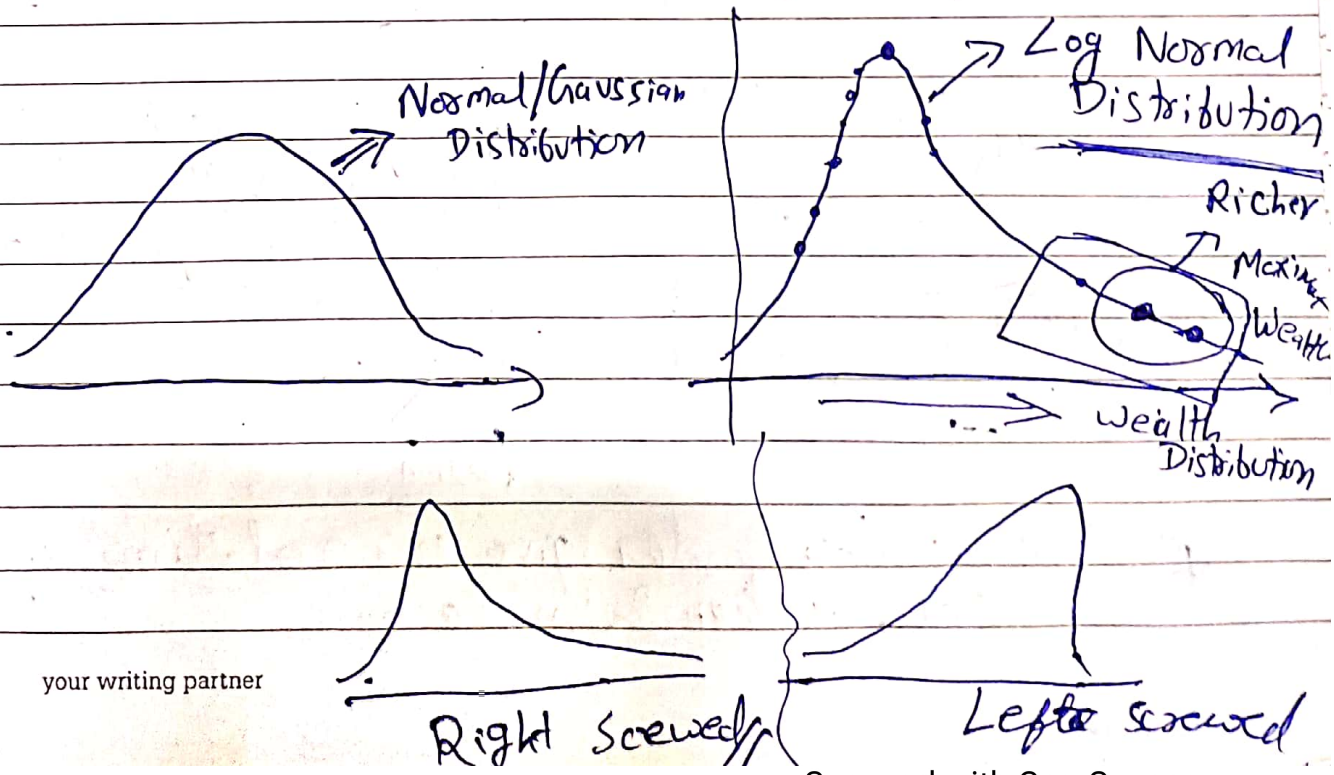
(a) Min Max Scaler ✓ ⟺ Standardization

ML ML



Min Max Scaler

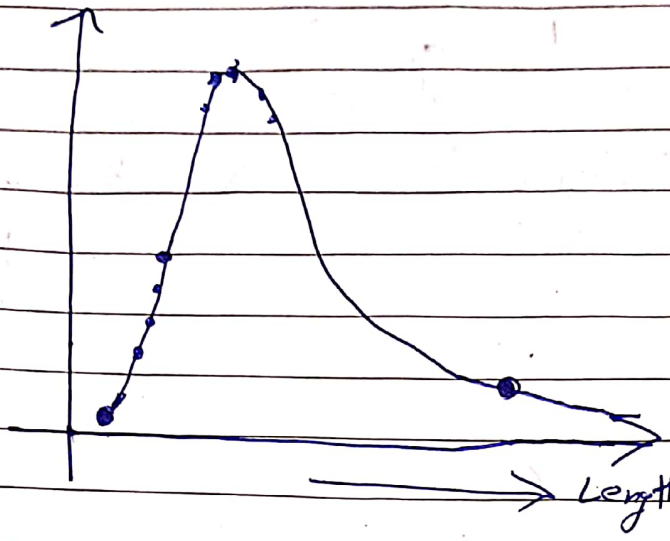
(A) Log Normal Distribution :-



your writing partner

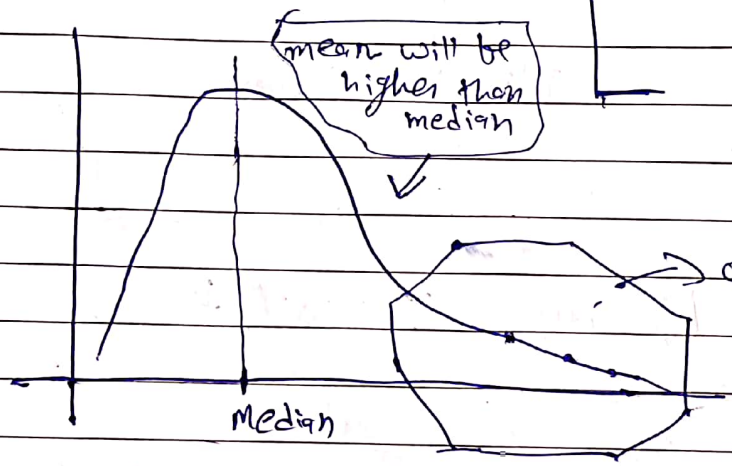
Subject _____

MON TUE WED THR FRI SAT SUN

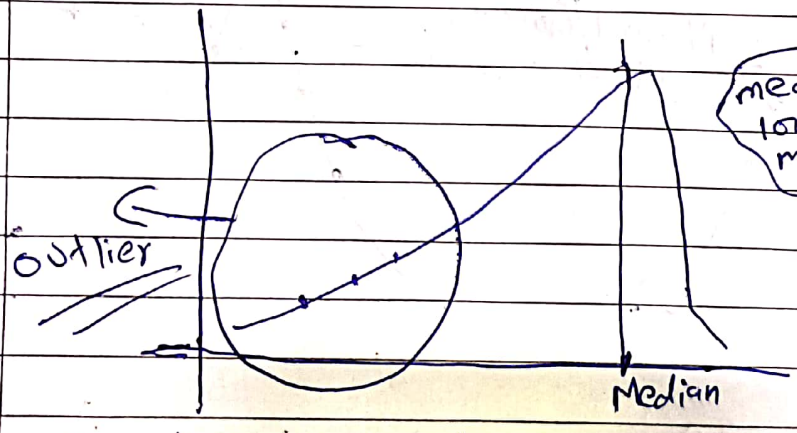


Assignment

Relation of Mean, Median, Mode



Graph 1



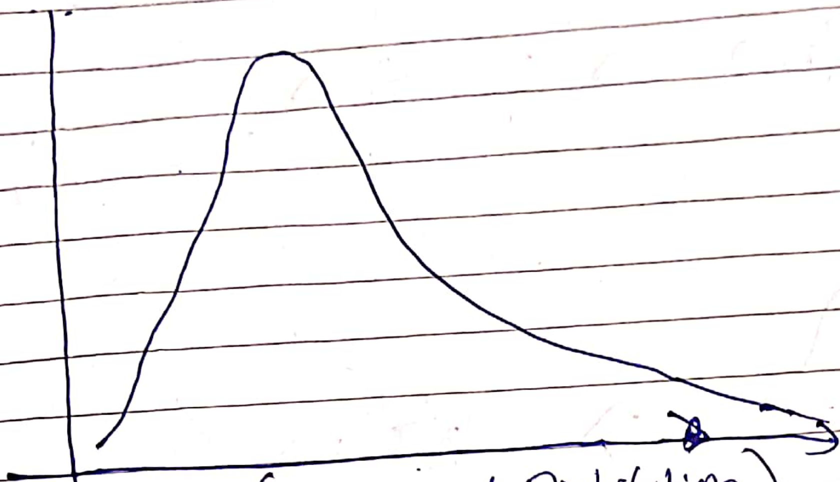
Graph 2

From Ascending order give the relation of mean, median & mode ??

① Log Normal Distribution is only

Subject

MON TUE WED THR FRI SAT SUN



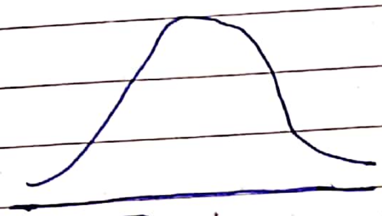
(Log Normal Distribution Graph)

$X \approx$ Log Normal Distribution

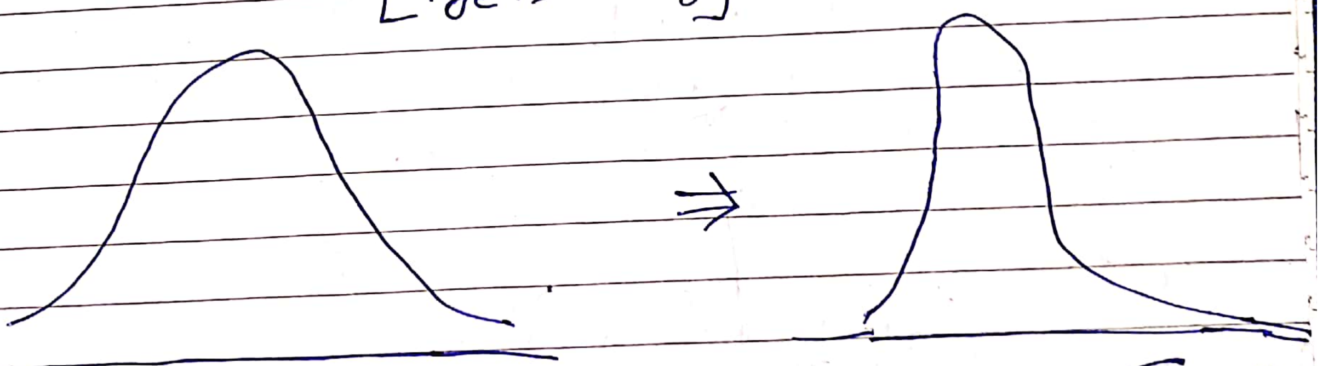
$y = \ln(x)$

[$\ln$ = log with the Base e (Natural Log)]

[$\log_e \Rightarrow$ Natural Log]



[Normal Distribution [Gaussian Dist]]



$X \approx ND(\mu, \sigma) \Rightarrow \exp(x)$ [Anti Log]

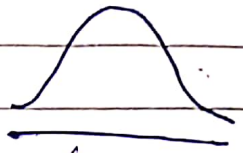
$\Downarrow$

y

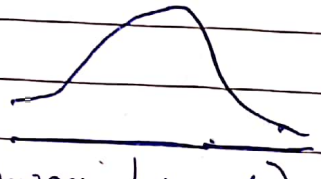
Subject _____

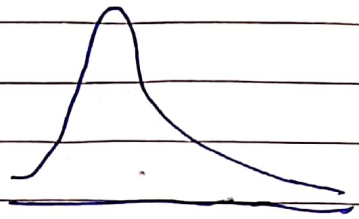
MON TUE WED THR FRI SAT SUN

Some Questions

Q.1 $X \sim LND \Rightarrow \log_e(x) \Rightarrow$  $\Rightarrow$ QQ Plot

(Gaussian/Normal Distribution)

Q.2 

(Gaussian/Normal Distribution) $\Rightarrow \exp(x) \Rightarrow$ 

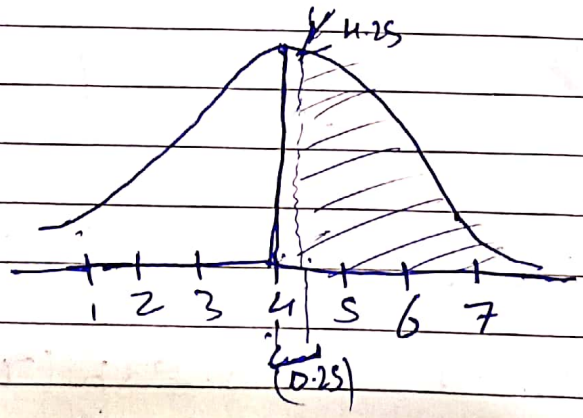
(Log Normal Distribution)

Problem Set

Eg: $X = \{1, 2, 3, 4, 5, 6, 7\}$

$\mu = 4$

$\sigma = 1$



Quest

your writing partner

www. z-table.com

Subject _____ MON TUE WED THR FRI SAT SUN

Question

① What is the percentage of score that falls above 4.25 ?

[where ⇒ μ = 4, σ = 1]

[Area of entire curve = 1]

① Z-score = $\frac{X - \mu}{\sigma} = \frac{4.25 - 4}{1} = 0.25$

⇒ 0.59 ⇒ 59%

② Z-table · (Area under the curve) [www.z-table.

⇒ 0.25 ⇒ 0.59 ⇒ 59% //

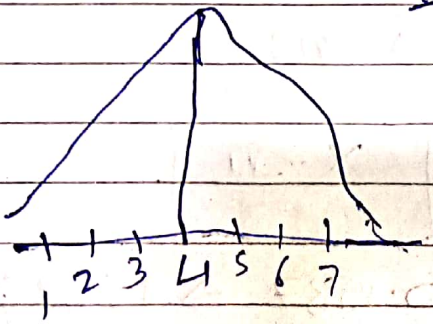
~~⇒ 0.59 ⇒ 59%~~

~~⇒ 1 - 0.59 ⇒ 0.41~~

⇒ 41%

② What is the percentage of score that falls below 3.75?

Soln Then, $\frac{3.75 - 4}{1}$ = ~~3.75~~ = 0.25



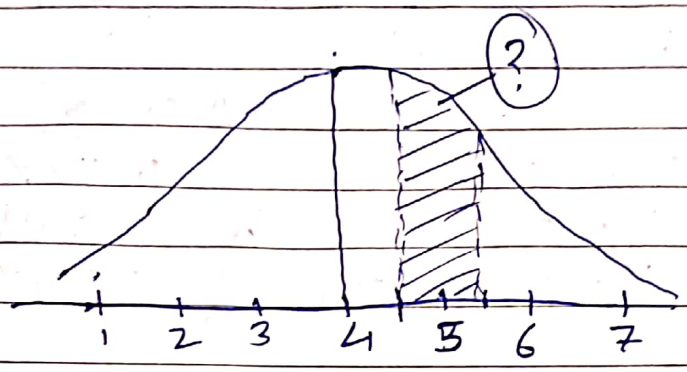
-0.25 ⇒ 40%

Approx Area = 40%

Subject _____ Problem Set

MON TUE WED THR FRI SAT SUN

Q) What is the percentage of score that falls between 4.75 & 5.75?



Given

$\mu = 4$

$\sigma = 1$

$Z\text{-score} \Rightarrow \frac{X - \mu}{\sigma}$

Now,

[Got By z-table]

for 5.75 $\Rightarrow \frac{5.75 - 4}{1} = 1.75 = 0.9599$

for 4.75 $\Rightarrow \frac{4.75 - 4}{1} = 0.75 = 0.7734$

Now

Value(5.75 - 4.75)

$\Rightarrow (0.9599 - 0.7734)$

$\Rightarrow 0.1865$

Approx Area $\Rightarrow$ 18.65 %

Subject _____ Problem Set

MON TUE WED THR FRI SAT SUN

Q) In India the Average IQ is 100 with a Standard Deviation of 15, What is the percentage of population would you except to have an IQ for :-

Answer
- (a) Lower than 85 — 0.1587
- (b) Higher than 85 — 0.8413
- (c) Between 85 and 100 — 0.3413

(A) Solution

Given,

$\mu = 100$

$\sigma = 15$

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

Lower than 85

(a) $\frac{85-100}{15} \Rightarrow \frac{-15}{15} \Rightarrow -1 \Rightarrow 0.1587 \Rightarrow 15.87\%$

Higher than 85

(b) $1 - 0.1587 \Rightarrow 0.8413 \Rightarrow 84.13\%$

Between 85 & 100

(c) for 100 $\Rightarrow \frac{100-100}{15} \Rightarrow 0 \Rightarrow 0.5$

for 85 $\Rightarrow \frac{85-100}{15} \Rightarrow \frac{-15}{15} \Rightarrow -1 \Rightarrow 0.1587$

(A) Between 85 & 100

$\Rightarrow 0.5 - 0.1587 = 0.3413$

$\Rightarrow 34.13\%$

your writing partner