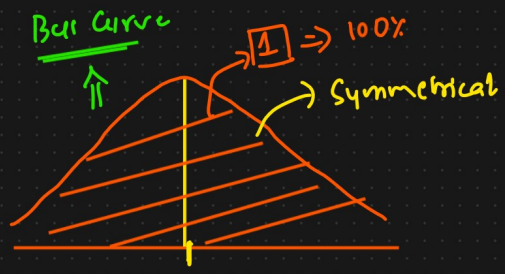
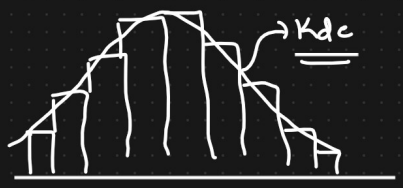
Day 3 - Stats

1. Normal Distribution ✓
2. Standard Normal Distribution ✓
3. Z-score ✓
4. Standardization And Normalization ✓
5. Gaussian / Normal Distribution



Kernel density estimator



Age, weight, height → Dismbution

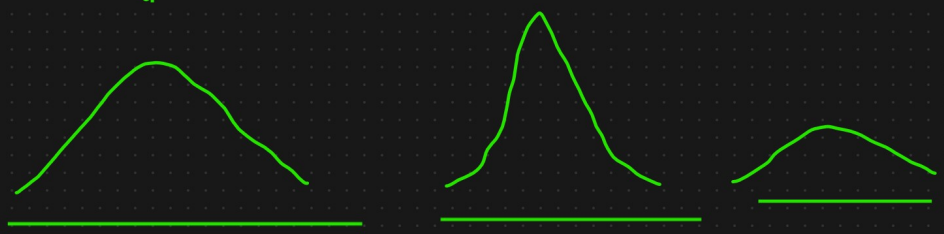
=> Doctors

Domain Expertise

IRIS DATASET ⇐

Petal length, Sepal length, petal width, Sepal width

Gaussian Distri



6. [Empirical Rule of Normal Distribution]

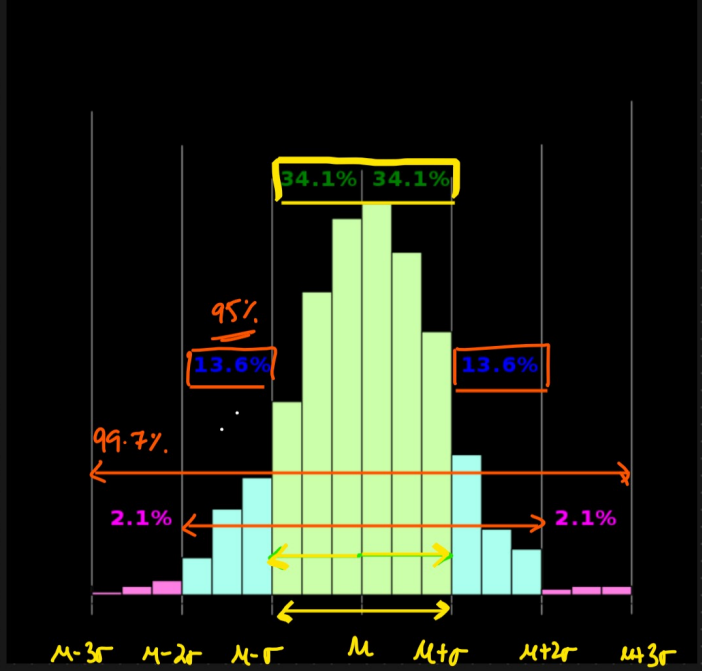
Empirical Rule

⇑

68 - 95 - 99.7 %

⇓

CLEAR



Gaussian / Normal Distribution

⇓

Assumptions of the data

Age = { } ⇒ Gaussian / Normal Dist

✓

Q-Q plot ⇒ Distribution is Gaussian or Not?

Standard Normal Distribution

$X \sim$ Gaussian Distribution (μ, σ)

⇓

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

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

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

⇓

$\mu = 3$

$\sigma = 1.41$

$Z\text{-score} = \frac{x_i - \mu}{\frac{\sigma}{\sqrt{n}}}$ $\quad n = 1$

⇒ Standard Error ⇒ Inferential stats.

$Z\text{-score} = \left| \frac{x_i - \mu}{\sigma} \right|$ ⇐ Simple

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

$\mu = 3 \quad \sigma = 1.41$ &

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

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

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

$\rightarrow Y = \{-1.414, -0.707, 0, 0.707, 1.414\}$

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

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

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

0-1

Same Scale

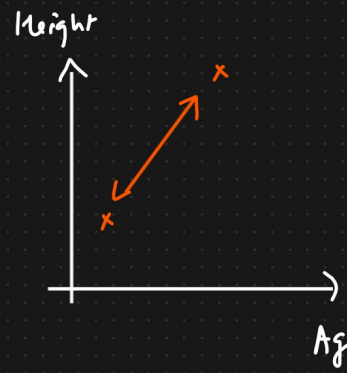
Machine Learning

Maths Equations

$\Downarrow$

Algorithm $\Rightarrow$ Mathematical Model

Mathematical Calculation Time $\uparrow\uparrow\uparrow$



Feature Scaling

✓ Normalization $\Rightarrow$

$\Downarrow$

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

$[0 - 5]$

$[0 - 4]$

Standardization $\{$ Z-score $\}$

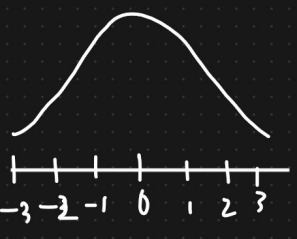
$\Downarrow$

$\mu = 0, \sigma = 1$

$[-3 \longleftrightarrow 3]$

$x_i - \mu$

$\Downarrow$



Normalization [lower scale ↔ Higher scale] → Images => 0-255

Standard

① Min Max Scaler [0-1]

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

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

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

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

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

x	y	y'
1	0	-1.414
2	0.25	-0.707
3	0.5	0
4	0.75	0.702
5	1	1.414

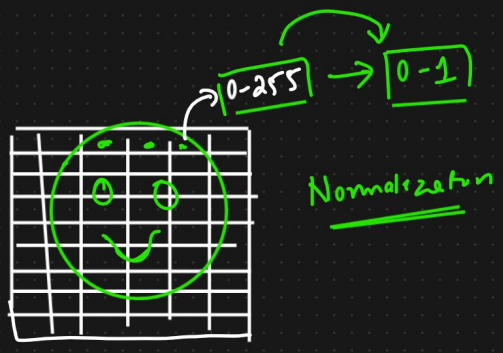
⇓ Apply ??

RNN ANN CNN

B.C PMD NASA

BF BF BF

Deep learning



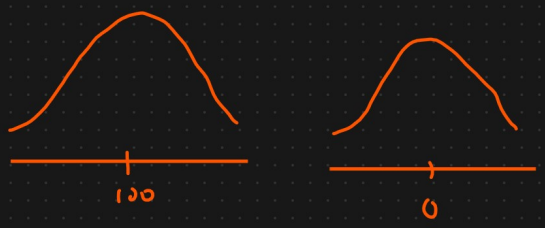
① Standardization

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

X → Normal Distribution (μ, σ)

⇓ Z-score.

y → SND $(\mu = 0, \sigma = 1)$



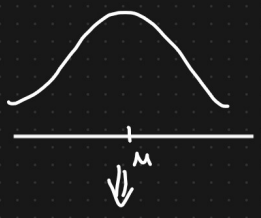
Why do we do this → Bring the features in the same scale

Normalization [0 - 1]

① Min Max Scaler ✓ ⇔ Standardization

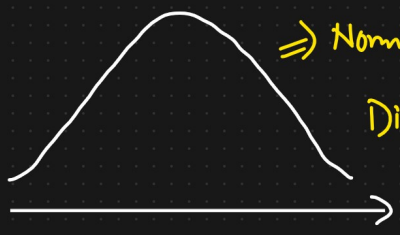
↓ ML

↓ ML



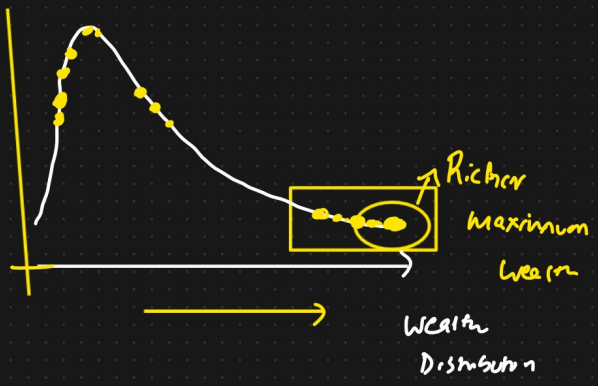
Min Max Scaler

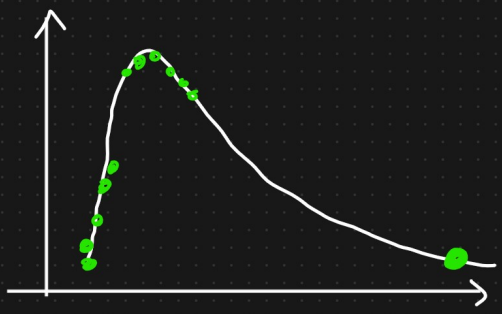
① Log Normal Distribution



⇒ Normal/Gaussian Distribution.

Log Normal Distribution



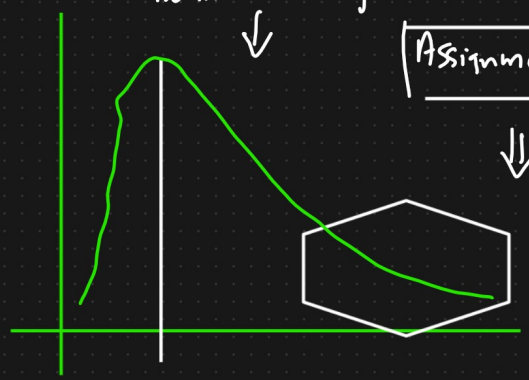


mean will be higher

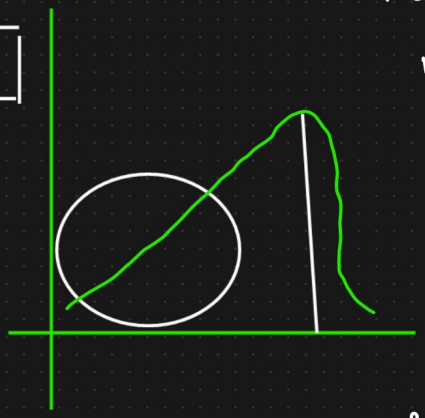
↓

Assignment

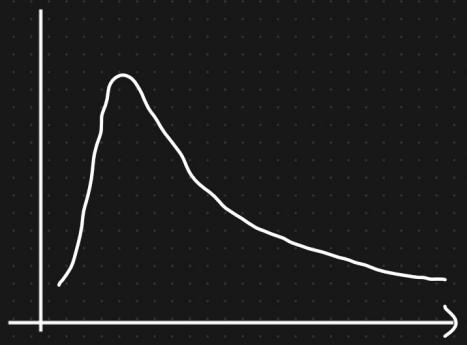
↓



Relation of mean, median, mode

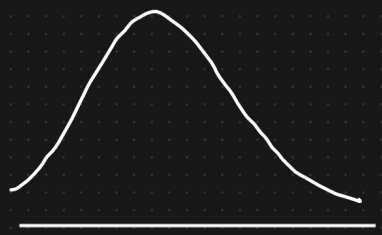


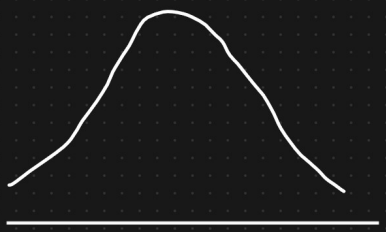
From Ascending order give the relation of mean, median & mode?



$X \sim$ Log Normal Distribution

$y = \ln(x)$



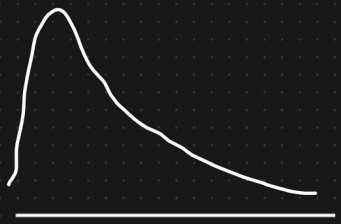


$X \sim ND(\mu, \sigma)$

antilog

$\Rightarrow \exp(x)$

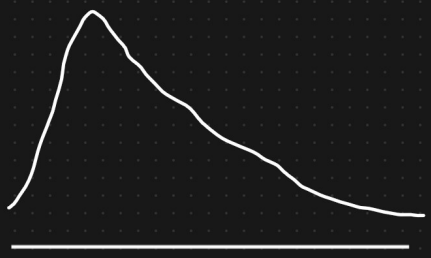
y



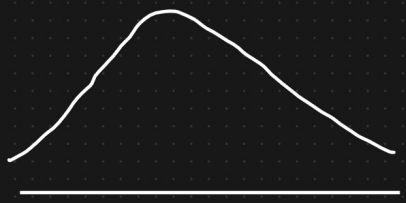
(¥)

Natural log

$\log_e$



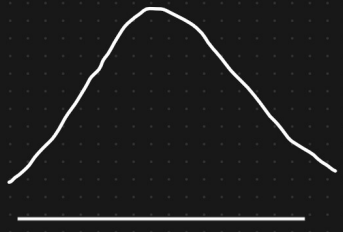
$\Rightarrow y = \ln(x)$



$X \sim$ Log Normal Distribution

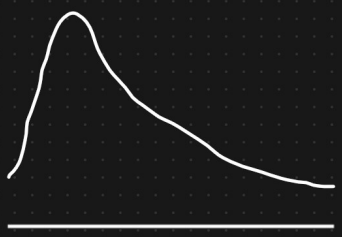
(μ, σ)

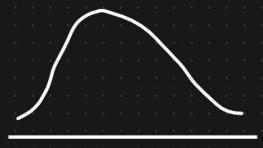
Inverse



y

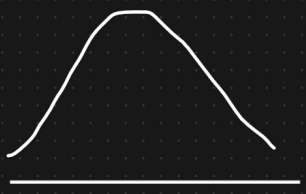
$\Rightarrow X = \exp(y)$

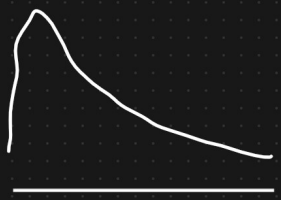


$X \sim f_{ND}$ $\Rightarrow$ $\log_c(x)$ $\Rightarrow$  $\Leftarrow$ QQ plot

$\Downarrow$

Tomorrow



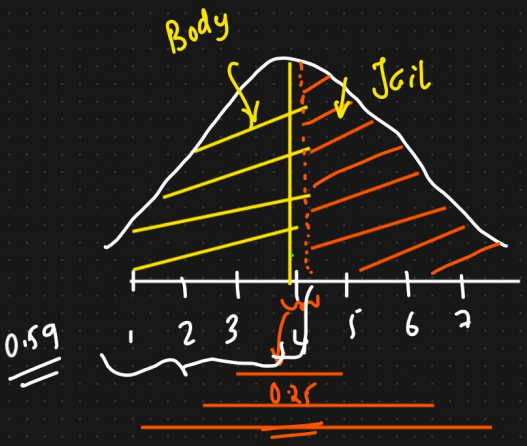
$\Rightarrow y = \exp(x)$ $\Rightarrow$ 

x

(#) $X = \{1, 2, 3, 4, 5, 6, 7\}$

$\mu = 4$

$\sigma = 1$



Question: What is the percentage of score that falls above 4.25?

fall below 3.75?

0.59 $\Rightarrow$ 59%

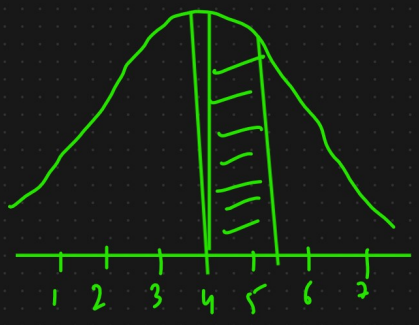
$1 - 0.59 = 0.41 \Rightarrow$ 41%

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

① Z-table (Area under the curve)

↳

$Z\text{-score} = \frac{3.75 - 4}{1} = -0.25 \approx 40\%$



4.75 & 5.25

(4) In India the average IQ is 100 with a Standard Deviation of 15. What is the percentage of population would you expect to have an IQ

Answers.

① Lower than 85 = 0.1587
② Higher than 85 = 0.8413
③ Between 85 and 100 = 0.3413

Assignment