Time Series

Non time Series

House Price Prediction

Regression

Size	location	Bedrooms	Price

$\boxed{Y = mx + c}$ → linear eq

$$\boxed{Y = m_1x_1 + m_2x_2 + m_3x_3 + \dots m_nx_n + c}$$

~~Time~~

time Series

hour, min, sec, Day, month, year → timestamp

Sales

Corr

Day 1 → 50k

Day 2 → 60k

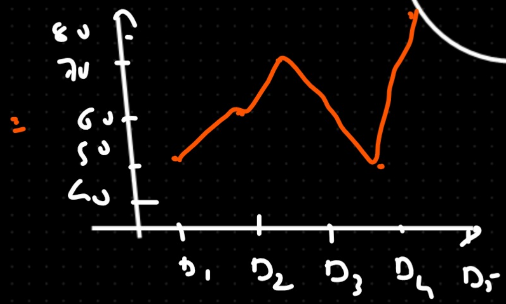
Day 3 → 70k

Day 4 → 40k

Days → 75k

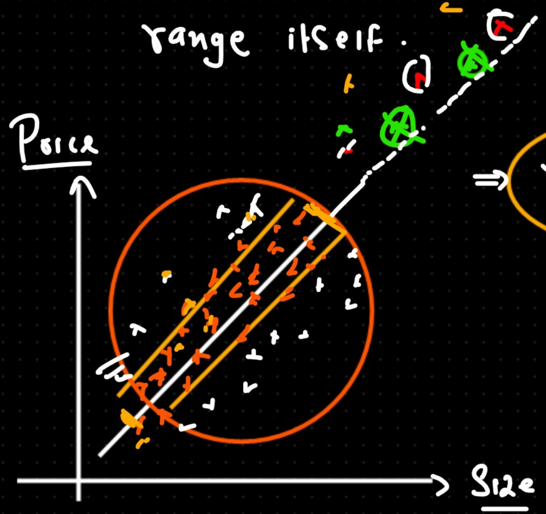
iNeuron → Sales

Mangment



Ques Can we solve this TS with linear regression?

Interpolation => to find out the value in the range itself.



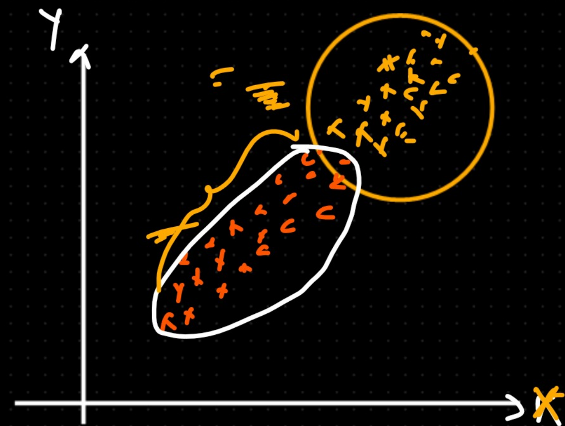
=> $Y = mx + c$

Orange training Data we have in the Particular range

test data

Non-time series Problem Data

- Extrapolation => to find out the value out of Range.



Forcasting

timeseries data

extrapolation

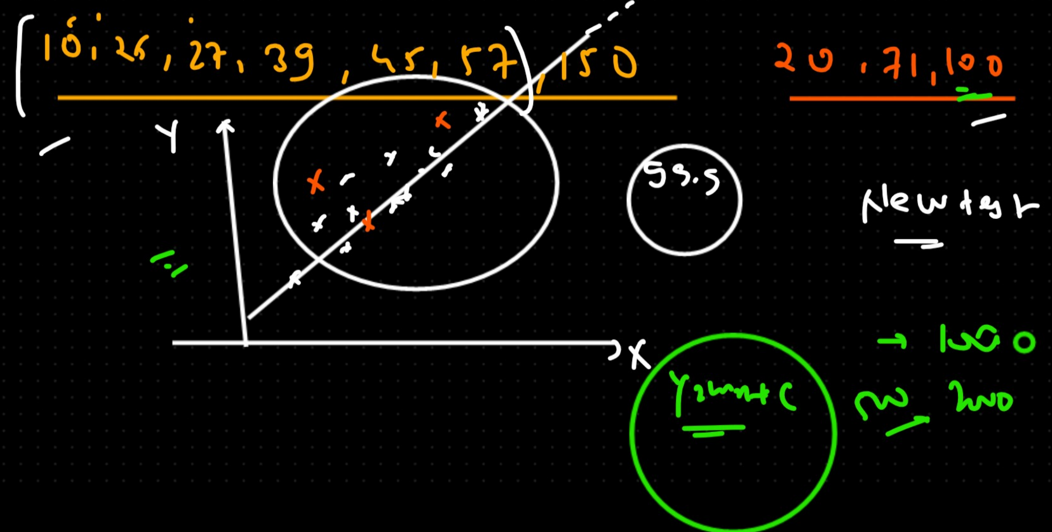
time series Problem statement will be extrapolation

- Based on Previous history forcast the future value.

Ques Can we solve this TS with linear regression?

1. Because of entrapolation. we can not do.
2. Because of outlier we can not use.
3. TS Data is more complex
4. Linear reg. assume that there should linear realtionship but in time you won't get linear realtionship
5. In Non-TS Data there is no such effect of time but TS - Data there is effect of time mean current TS is Dependent on prev. time stamp.

Data ⇒ [10, 20, 25, 27, 39, 45, 57, 71, 100, 150]



Example

1. economics forcasting ⇒ GDP, intrest rate, inflation rate
2. Finance → sales, Bond Price
3. Weather forcasting ⇒ Pattern of Diff season, Prediction of weather
4. Medical ⇒ Based on Prev. history of patiance need to predict future condition.

Time Series ⇒ time dependent data ⇒ any domain

- Data → Time dependent Data
- EDA → –
- Preprocessing → –
- model →
- model evaluation

Classical / Statical

- ARIMA
- SARIMA
- SARIMA X

⇒ S + AR + I + MA + X

equation ← Mathmatical

↓

forcaste

Deep learning Based approach

NLP

keeping a previous context from the data

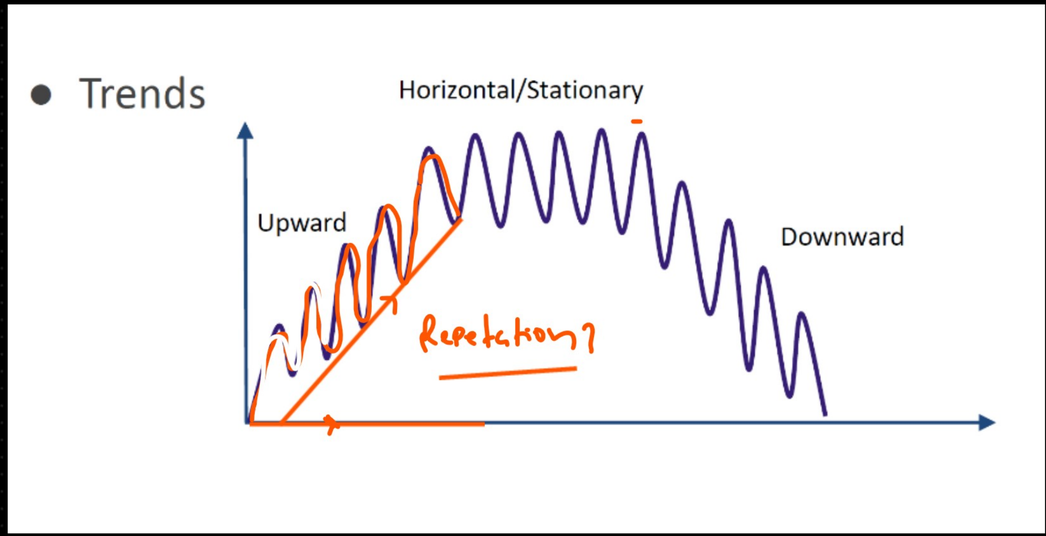
- RNN
- LSTM
- Attention
- transformer

⇒ Predict forcast

Readymade model

- N-beats
- Prophet

Timeseries ⇒ (1) trend
(2) Season
(3) Noise
(4) Cycle



Trend ⇒ (1) upward trend
(2) Downward trend
(3) flat (horizontal)

Season ⇒ Frequent (repetation) (Daily, monthly, yearly)

(1) Sales of the ice-creame in summar
(2) traffic situation at 6 pm in my area
(3) tourest in goa at the end of year (Dec, Jan)

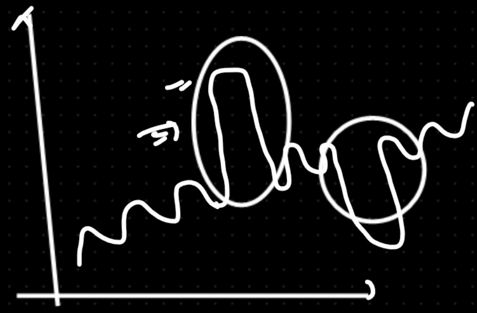
Noise ⇒ Some uncertainity or some randomness in my data. bcoz of unpredictable reason.

UNPREDICTABLE REASON ⇒

Pandamic ⇒ corona

War ⇒ R|U

Report ⇒ hindenberg report



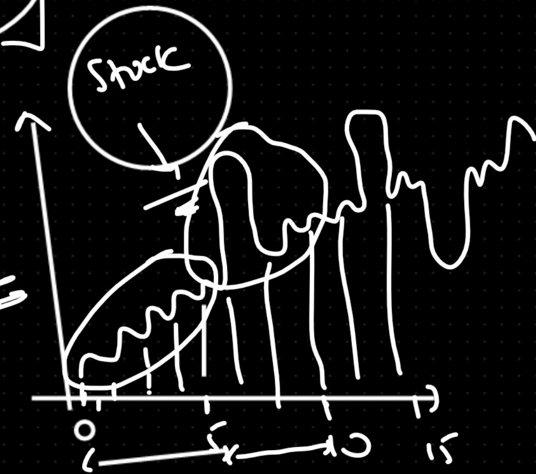
Cycle ⇒ time Series behaviour over the long Period of time

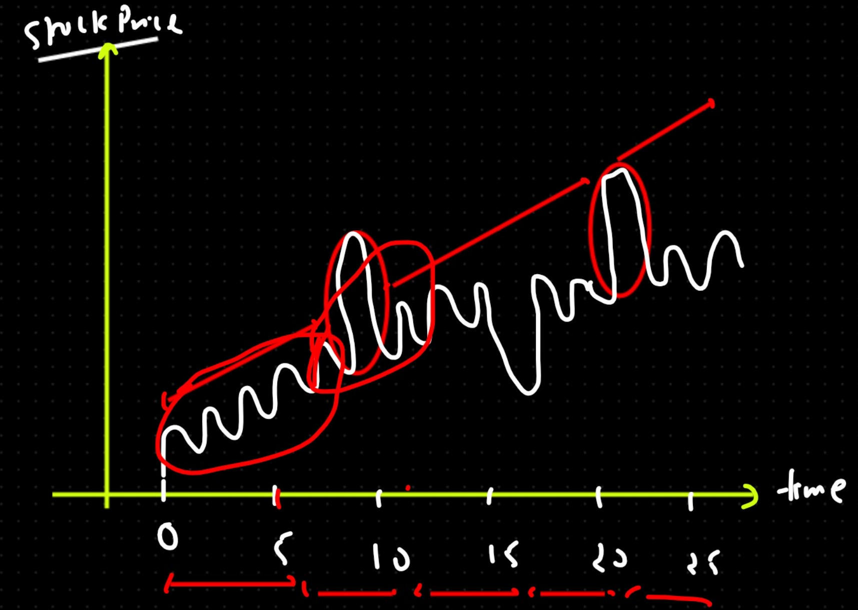
$\boxed{\text{Cycle} \Rightarrow \text{seasonality} + \text{Noise(flucation)}}$

Election ⇒ 5-Year ⇒ cycle

Cencus Counting ⇒ 10 year

economics ⇒ GDP





trend → upward
Seasonality ⇒ yes
⇒ Non stationary
Noise ⇒ yes

Cycle → long period

↓

Seasonal effect + Noise

Component ⇒ Multiplicative ⇒ $y_t = T \times S \times N$

1. Non-linear Pattern
2. Non_constant var

D1	50k
D2	60k
D3	70k
D4	80k
D5	50k
D6	100k

Additive ⇒ $y_t = T + S + N$

1. linear Pattern
2. It is having constant var.

Moving average

1. Simple moving average (SMA)
2. Commulative moving average (CMA)
3. EMA or EWMA

SMA =>

Simple moving average

Average of [10, 12, 15, 13, 11]

$$\frac{10+12+15+13+11}{5}$$

= 12.5 Approx

Moving average =>

Why we Calculate MA?

SMOOTHING Of TIME SERIES DATA

Moving => move over the time axis in given window size

Moving + average

Window Size → calculating average

Window Size = 3

D_1	10	NAN
D_2	12	NAN
D_3	15	→ 12.33
D_4	13	→ 13.33
D_5	14	→ 14
D_6	16	→ 14.33
D_7	17	→ 15.66

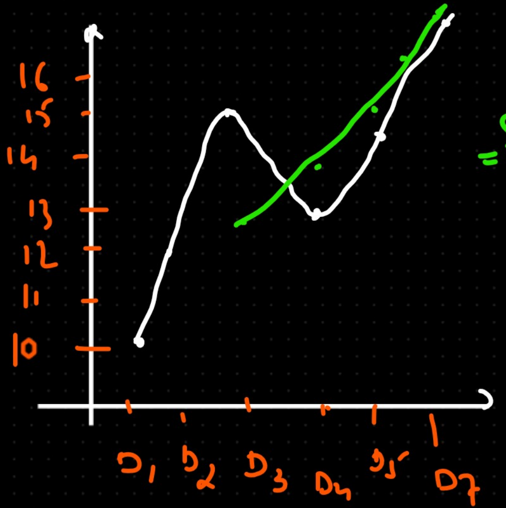
1st avg $= \dfrac{D_1 + D_2 + D_3}{3} = \dfrac{10 + 12 + 15}{3} = 12.33$

2nd avg $= \dfrac{D_2 + D_3 + D_4}{3} = \dfrac{12 + 15 + 13}{3} = 13.33$

3rd avg $= \dfrac{D_3 + D_4 + D_5}{3} = \dfrac{15 + 13 + 14}{3} = 14$

4th $= \dfrac{D_4 + D_5 + D_6}{3} = \dfrac{13 + 14 + 16}{3} = 14.33$

5th $= \dfrac{D_5 + D_6 + D_7}{3} = \dfrac{14 + 16 + 17}{3} = 15.66$



= Smoothing ?

T x S x N

1. to remove all the effect from the data
2. Pattren recognization from the data
3. you can easily analysis trend of Data
4. you can reduce the effect of outlier
5. you can easily visulizing the data.

CMA => Find out the avg of all the Data Points up to the given time stamp

D_1	10	=>	$D_1 = 10$
D_2	12	=>	$\frac{D_1 + D_2}{2} = 10+12/2 =$
D_3	15	=>	$\frac{D_1 + D_2 + D_3}{3}$
D_4	14	=>	$\frac{D_1 + D_2 + D_3 + D_4}{4}$
D_5	16	=>	$D_1 + \dots D_5/5$
D_6	17	=>	$D_1 + D_2 + \dots D_6/6$
D_7	18	=>	$D_1 + D_2 + D_3 \dots D_7/7$

$$\frac{D_1 + D_2 + D_3 + \dots + D_n}{n}$$

① It will give you the exponential trend.

② We use CMA for the long time period.

EMA or EWMA ⇒ in EMA we give more weightage to the recent Data point or give more weightage to the recent time stamp.

$$\boxed{V_t = \beta V_{t-1} + (1-\beta)\theta_t}$$

V_t = EMA at time t

β = $0 < \beta < 1$ ⇒ 0.9

V_{t-1} = EMA at Previous time stamp

θ_t = Data value at t time stamp

			EMA
V_0 →	D_1	25	⇒ 0
V_1 →	D_2	13	θ_1 ⇒ 1.3
V_2 →	D_3	17	θ_2 ⇒ 2.8
V_3 →	D_4	31	θ_3 ⇒
V_4 →	D_5	43	θ_4 ⇒

$V_0 = 0$ or $V_0 = 25$

$V_1 = \beta \times V_0 + (1-\beta)\theta_1$

$= 0.9 \times 0 + (1-0.9) \times 13$

$= 0 + 0.1 \times 13 = 1.3$

$V_2 = \beta \times V_1 + (1-\beta)\theta_2$

$= 0.9 \times 1.3 + (1-0.9) \times 17$

$= 1.17 + (0.1 \times 17)$

$= 1.17 + 1.7$

$= 2.8$

1st 2nd 3rd 4th 5th 10th 11th 12th

EMA

D1 15k
20k
25k
30k
35k

1st 2nd

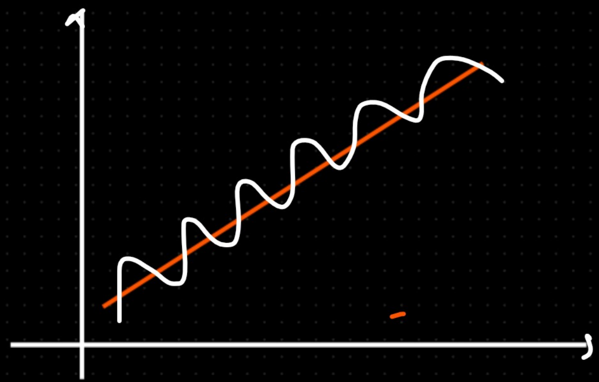
10m — 11m — 12th

DL

Gradient

chain rule

Stationary + Non-stationary



trend = upward

MA (SMA)

window + average

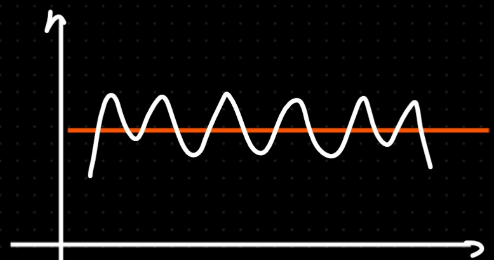
=) my moving is not constant over the time.

=) variance is not constant over the time.

Non-stationary = Mean and var will not be const

Stationary = Mean and var will be constant

{the meaning is there is no change in Mean, var over the time axis}



trend is horizontal

make it stationary ?

Data → EDA → PreProcessing → Model

ARIMA, DL

Non-stationary ?

Analysis ⇒ 1. Visulization
2. Stats based test

Preprocess ⇒ (Non-st to st) ⇒ 1. Differencing
2. log
3. root
4. Adjustment of Seasonal Data

ACF and PACF and ARIMA

ACF → auto corr fln

PACF ⇒ Partial auto corr fln

ACF ⇒ Auto + Correlation

Correlation itself in the Feature

It is a relationship B/w two feature

x_1 | x_2

- Pearson
- Spearman Rank
- Kendall

ACF Measure the Corr B/w time Series and it lag value

		1st lag	2nd lag	3rd lag
D1	10	NA	NA	NA
D2	25	D1 — 10	NA	NA
D3	14	D2 — 25	D1	NA
D4	$\boxed{16}$	D3 — 14	D2	D1
D5	20	D4 — 16	D3	D2
D6	32	D5 — 20	D4	D3

y_t ↓ y_{t-1} x_1 x_2 y_{t-2} y_{t-3}

D_2 D_1

D_3 D_2

D_4 D_3

D_5 D_4

D_6 D_5

25	10
14	25
16	14
20	16
32	20

⇒ Corr itself in the feature | Autocorr with lag 1

Corr with 1st lag

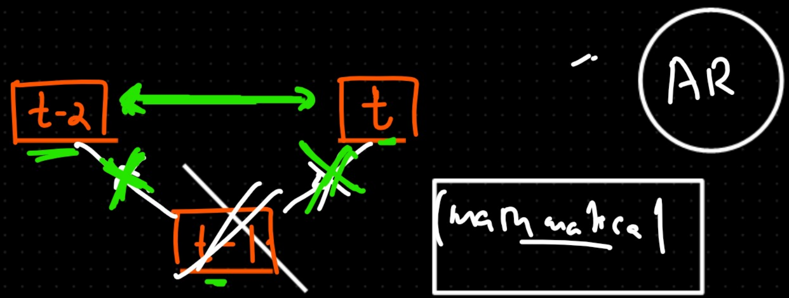
→ $\text{Corr}(y_t, y_{t-1})$

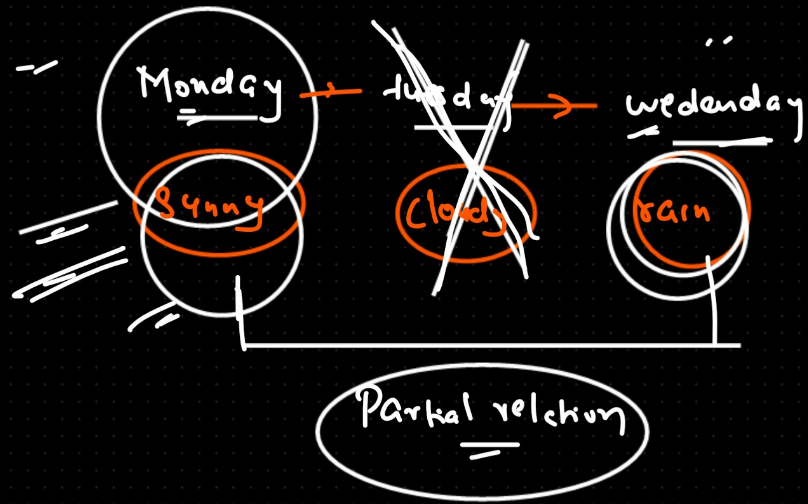
Corr with 2nd lag

$\text{Corr}(y_t, y_{t-2})$

= $\text{Corr}(y_t, y_{t-3})$

PACF => Partial auto corr





Auto regression

Auto Regression

AR

{regression itself in the var}

$x \rightarrow y$

$x \rightarrow$ independent feature

$y \rightarrow$ dependent feat

$y = mx + c$

m = slope

c = intercept

$\boxed{y = m_1x_1 + m_2x_2 + m_3x_3 \ldots m_nx_n + c}$

ϕ

- y_t = value at the current time stamp
- ψ = coff term
- C = constant
- ε = error

AR eq. with lag 1 $\quad \boxed{y_t = \psi y_{t-1} + C}$

AR eq. with lag 2 $\quad \boxed{y_t = \psi_1 y_{t-1} + \psi_2 y_{t-2} + c}$

AR eq. with lag n $\quad \boxed{y_t = \psi_1 y_{t-1} + \psi_2 y_{t-2} + \cdots + \psi_n y_{t-n} + c}$

AR + I + MA → MAM

star