* Time series → Time component is present in the data.

*

Time Series

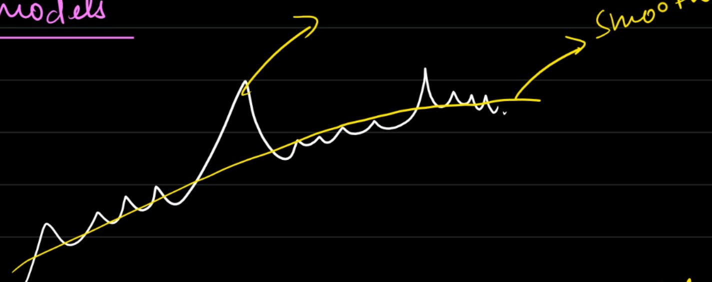
- Time series Analysis
  - Level
  - Trend ✓
  - Seasonality ✓
  - Cyclicity
  - Noise ✓
- Time Series Forecasting
  - Smoothening models
  - Forecasting models

Trough & Crest | rough

* Smoothening models

⇓

Smoothing



Smoothing of time series → removing fluctuations.

1′ → Naive Model → last observed value is the forecast.
1″ → Average Model → Average value is the prediction.

1. Simple moving Avg (SMA)
2. Cumulative moving Avg (CMA)
3. Exponential weighted Moving Average (EWMA)

① SMA.

$$\text{Avg} = \frac{\text{Sum of all nos}}{\text{no. of values}}$$

$$2, 3, 4, 5 \Rightarrow \frac{2+3+4+5}{4} = \frac{14}{4} = 3.5$$

* MA ⇒ moves over the time axis in a specific window. (window size, Avg value)

Window 3
Avg.

		SMA=3
Jan	50	Na
Feb	65	Na.
Mar	70	62
April	85	74
May	90	82
June	100	80
July	110	90
Aug		95
Sep		L
Oct		

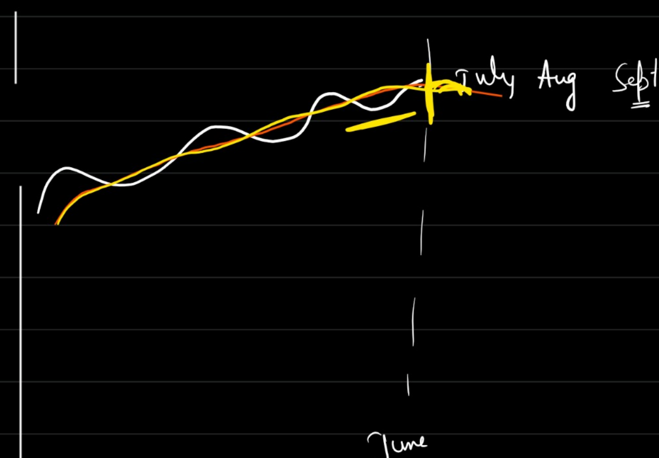
1st avg $= \frac{50+65+70}{3} = 62$

2nd avg $= \frac{65+70+85}{3} = 74$

3rd avg $= \frac{70+85+90}{3} = 82$

W.S = 2

1	Na
2	$\frac{1+2}{2}$
3	$\frac{2+3}{2}$
4	$\frac{3+4}{2}$
5	$\frac{5+4}{2}$



* Why smoothiy

→ To check trend of date.

→ To remone all the effect from date (Fluctuation)

→ redue the effect of oultler

→ visualisation

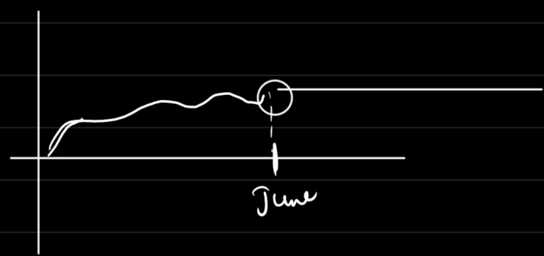
② Cumulative Moving Average

→ find all the avg of all the dp's upto given timestamp.

		CMA
Jan	10	10
Feb	12	$\frac{10+12}{2}$
Mar	15	$\frac{10+12+15}{3}$
Apri	14	$\frac{10+12+15+14}{4}$

→ for long period of time.

→ Exponential trend



③ EMA or EWMA

SMA, CMA, Naive, Avg

→ It gives equal priority to all the values.

→ In time series current observation is highly influenced by last few observation.
∴ You need to give priority to recent observation

⇓

VK.

1 0 0 1 0 1 | 76 |

→ We give more weightage / Importance / priority to the recent datapoint / time stamp

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

V_t = EMA at time t

β = $0 < \beta < 1$ → generally its 0.9 (weight)

V_{t-1} = EMA at previous ts.

θ_t – Data at current stamp

$\beta = 0.9$ } To reduce the prev time stamp effect → $\frac{1}{1-\text{Beta}}$

V_{t-1}	Month	Value	EMA
V_0	D_1	25	0
V_1	D_2	13	1.3
V_2	D_3	17	2.87
V_3	D_4	31	–
V_4	D_5	43	–

$V_1 = \beta \cdot V_0 + (1-\beta)\theta$

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

$= 0.1 \times 13 \Rightarrow 1.3$

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

$= 0.9 \times 1.3 + 0.1 \times 17$

$= 2.87$

time Series

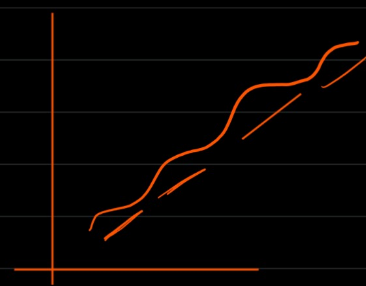
- Additive
- Multiplicative

Additive

y_t = Trend + Season + Noise

→ Linear over time
→ Constance variance
→ increased trend at same difference

Day 1	100
Day 2	200
Day 3	300
Day 4	400



Multiplicative

y_t = Trend x Seasonality x Noise

→ Non linear
→ Non constant variance

price

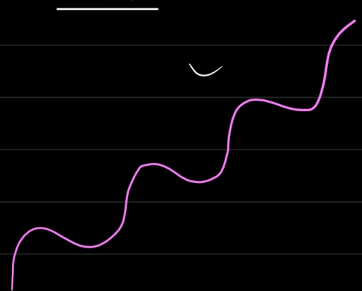
Time series

- Analysis
- forecasting
  - Smoothing
  - Autoregressive models.

→ you can build a smothening time series model on any time series data

* To build Autoregressive model, statistical properties of a time series like mean / variance should bee constant (not change over time)

Scenario 1



Scenario-2

→ It will be easy to build a forecasting model in SC-2
→ less variance in data

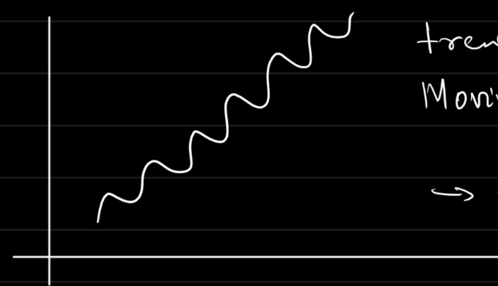
Non stationarity → Mean - variance will not be constant.
Stationarity → over the time, value is not varying (changing)

* To build a Autoregressive model, make the time series stationary.

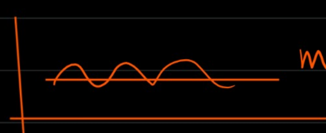
* ML — Data ingestion → Analysis → preprocessing — Model building
* Time series → " → TSA → Timeseries is stationary → build forecasting model
↳ if nonstation → convert to stationary ts

* To check if a TS is stationary or not

① Visualisation



trend = upward
Moving avg is increasing over time
→ Variance is not const.
⇓
TS is non stationary.



moving avg const → timeseries is stationary.

② statistical test

ⓐ ADF (Augmented dickey fuller test)

H_0 : TS data is non station
H_A : TS " is stationary

p value ≤ 0.05 → reject H_0
Conclude - TS data is stationary

ⓑ KPSS test (Kwiatkowski - Philips - Schmidt - Shin test)

H_0 : TS data is stationary

* How to Convert non-st. time series to stationary ts

① Differencing
② log transformation — take the log of value
③ root " → " " sq root of values
④ Seasonality adjustment
⑤ Box-Cox transformation (to make variance constant → +ve)
⑥ Yeo-Jhonson transfo. (All ts data)

Differencing → Difference ($y_t - y_{t-1}$)

(current) Previous

Month	Price	1st diff	2nd order	3rd
J	5	NA	Na	
F	10	10-5=5	Na	
M	6	6-10=-4	-4-(5)=-1	
A	8	8-6=2	2-(-4)=6	
M	15	7		
J	7	-8		

→ Check St → 3rd order

→ After differencing check if TS is stationary

Statistical test (ADF) | Visualisation (using MA)

* if stationary → build the forecasting model
else → Again do differencing.

* ACF → Auto Correlation Function
* PACF → Partial Autocorrelation function
* Autoregression.

* ACF → Auto + Correlation

Auto ↓ Correlation itself in the feature.

Correlation ↓ Relationship b/w two variables.

* ACF measures the correlation b/w time series & its lag value.

Month	y_t	1st lag	2nd lag	3rd lag
Jan	10	NA	NA	Na
Feb	25	10 (Jan)	Na	Na
Mar	35	25 (Feb)	10 Jan	Na
APR	42	35	25 Feb	10 Jan
May	50	42	35 Mar	25
June	55	50	42	35
July	62	55	50	42
	y_t	y_{t-1}	y_{t-2}	y_{t-3}

$Corr(y_t, y_{t-1})$

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

$Cor(y_t, y_{t-3})$

* Partial Autocorrelation Function

$t-2 \leftarrow \cancel{t-1} \leftarrow t$

remove the intermediate effect

	Month	y_t	$y_t - y_{t-2}$
t-6	Jan	10	
t-5	F	18 ×	
t-4	M	25	15 (25-10)
t-3	A	29 ×	
t-2	May	35	10 (35-25)
t-1	J	45 ×	
t	July	50	15 (50-35)

25	15
35	10
50	15

↓

PACF

Monday ~~tuesday~~ wed

Sunny ~~cloudy~~ raining

Autoregression

Auto — regression

regression with itself.

$$y_t = \psi y_{t-1} + c$$

y_t - value at current ts

ψ - coeff

c - constant.

$$y_t = \psi t_{-1} + \psi_2 y_{t2} + c$$

$$\boxed{y_t = \psi_1 y_{t-1} + \psi_2 y_{t-2} + \psi_3 y_{t-3} + \dots \psi_n y_{t-n} + c}$$

X | y

X - IV

y - DV

$y = mx + c$

$y = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots \theta_n x_n$

Autoregressive models

- → AR
- → MA
- → ARIMA
- → SARIMA
- → SARIMAX

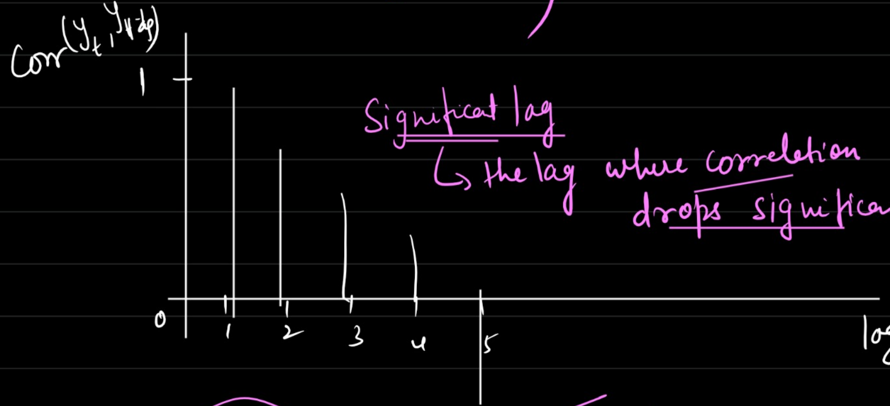
AR	I	MA
(Autoregression)	Integrated	(Moving Average)
⇓	⇓	⇓
p	d	q
$(0, 1, 2, \ldots n)$	$(0, 1, 2, \ldots n)$	$(0, 1, 2, \ldots n)$
lag value	lag value	lag value
⇓	⇓	⇓
PACF plot (correlogram)	Differencing (stationary)	ACF Plot (correlogram)

ACF (for MA) (q)

$Corr(y_t \; y_{t-1})$

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

$Corr(y_t \; y_{t-3})$



Significant lag

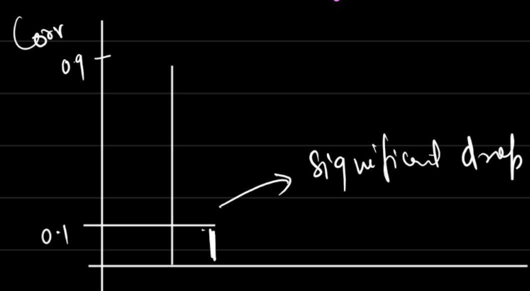
↳ the lag where correlation drops significantly.

Correllogram; lag 5

$q = 5$.

PACF

$y_{t-2} \rightarrow \cancel{y_{t-1}} - y_t$



→ significant drop – value of p in AR Model

MA – q (ACF)

AR – p (PACF)

(AR)

Autoregression

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \ldots \varphi_n y_{t-n} + c$$

Till how much lag we have to model?

⇓

PACF ⇒ p

(significant lag)

* we decide p (PACF) of AR, by seeing PACF plot and q (ACF) of MA by seeing ACF plot.

Integrated (differencing)

$y_t - y_{t-1}$ — D_1

$y_t - y_{y-2}$ — D_2

* MA. (Moving Average) (Not same as of smoothing)

q — ACF.

→ It models the error.

$$y_t = \varepsilon_{t-1}\varphi + c$$

(error)

$$y_t = \varepsilon_{t-1}\varphi_{t-1} + \varepsilon_{t-2}\varphi_{t-2} + \varepsilon_{t-3}\varphi_{t-3} + \dots \varepsilon_{t-n}\varphi_{t-n} + c$$

$$\text{ARIMA} = \underbrace{y_{t-1}\varphi_{t-1} + y_{t-2}\varphi_{t-2} + \dots y_{t-n}\varphi_{t-n}}_{\text{AR (p – PACF)}} + \underbrace{y_{t-1} - y_{t-2} - y_{t-3} \dots y_{t-n}}_{\text{Integral (differencing – d)}}$$

$$+ \underbrace{\varepsilon_{t-1}\varphi_{t-1} + \varepsilon_{t-2}\varphi_{t-2} + \dots \varepsilon_{t-n}\varphi_{t-n} + c}_{\text{MA (q – ACF)}}$$

AR (1)
I (2)
MA (1)

ARIMA (p, d, q)
(1 2 1)

$$y_t = \underbrace{y_{t-1}\varphi_{t-1}}_{\text{AR of order 1}} + \underbrace{y_t - y_{t-1} - y_{t-2}}_{\text{differencing of order 2}} + \underbrace{\varepsilon_{t-1}\varphi_{t-1}}_{\text{MA or order 1}} + c$$

p d q
↓ ↑ ↑
(1, 2, 1)

φ – Coefficient

y_t – obs

ε_t – Error

c – intercept.

SARIMA
↓
Seasonality

$$\text{ARIMA} = (p, d, q)\overbrace{(P, D, Q)}^{\text{season}}_{s}$$

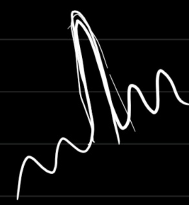
(1, 1, 2) $(1, 0, 2)_{12}$

→ Same p, d, q for Seasonal Component.

SARIMAX

↓ Seasonal

↓ Exogenous Variable (outlier, extra effect)



$(p\ d\ q)(P\ D\ Q)_s$ y_t (some extra variable)

Date	Sales	is_summer_vac
Jan	150	0
Feb	160	0
Mar	170	0
A	180	0
M		
J	250	1

Time series

↓ Smoothening model

→ Exponential smoothing

→ double Exponential Smoothing (Holt's method) — trend

→ triple " " (Holt winter method) → seasonality

Sources

→ Git - time series

→ Kaggle glossary.

→ Scikit cheatsheet.

Books

→ Probabilistic Models by Murphy

→ Pattern recognition by Christopher Bishop.

Stats

think stat

Extra topics

→ Cat Boost.

→ Light GBM

Dimensionality

→ LDA

→ t-SNE

USL

K-mode Clustering

Youtube Channel

- 3 Blue 1 Brown (ML, DL)
- Statquest (statistics)
- Python (Corey Schafer)

Certifications

→ Intro to ML (Andrew Ng Coursera)

→ Deeplearning Specialisation

→ deeplearning.ai

Stanford lecs