

ARIMA

ARIMA \Rightarrow Auto regression integrated moving average

Machine Learning

Data \rightarrow [model] \rightarrow mathematical

linreg
logreg
SVM
DT

time series data \rightarrow [model]
 \rightarrow classification \rightarrow ARIMA, SARIMA, SARIMAX
DL \sim RNN, Attention, transformation

AR
(auto regression)

\downarrow

P

$[0, 1, 2, 3, \dots, n] \rightarrow$ lag

\downarrow

PACF

I
= integrated

\downarrow

d (differencing)

$[0, 1, 2, 3, \dots, n] \rightarrow$ lag

(Non-stationary, Stationary)

MAIA
(moving average)

\downarrow

q

$[0, 1, 2, 3, \dots, n] \rightarrow$ lag

\downarrow

PACF

Auto regression

Auto

Regression

Regression itself
in the var

X	Y
-	-
-	-
-	-

X → independent
Y → dependent
|X→y|

timeseries lag1

D ₁	10	NAN
D ₂	20	D ₁ → 10
D ₃	30	D ₂ → 20
D ₄	40	D ₃ → 30
D ₅	50	D ₄ → 40

ACF

corr

↓ Y_{t-1}

20	10
30	120
40	130
50	140
?	50

?? ? } focus on

$$Y = mx + c$$

$$Y = m_1x_1 + m_2x_2 + m_3x_3 + \dots + m_nx_n + c$$

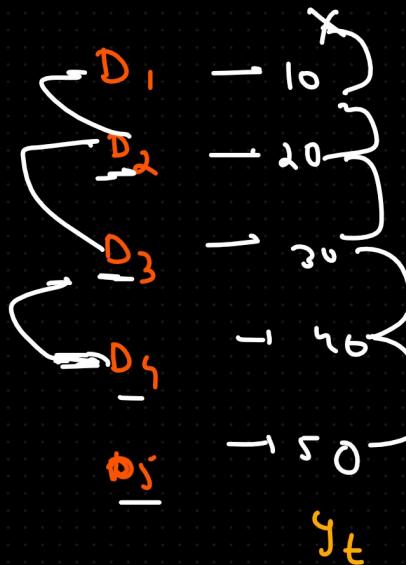
Autoreg
General eq.

$$y_t = y_{t-1}\psi + c \rightarrow \text{lag1}, p=1, \text{AR(1)}$$

$$\Rightarrow y_t = y_{t-1}\psi_1 + y_{t-2}\psi_2 + y_{t-3}\psi_3 + \dots + y_{t-h}\psi_h$$

Integration \Rightarrow Differencing

$I^1 + \text{level } D_1, t-1 \text{ lag}$



$$\begin{aligned} \text{NR} &= 10 \\ 20 - 10 &= 10 \\ 30 - 20 &= 10 \\ 40 - 30 &= 10 \\ 50 - 40 &= 10 \end{aligned}$$

$$\begin{aligned} \text{NAN} & \\ \text{NaN} & \\ 0 & \\ 0 & \\ 0 & \end{aligned}$$

$$\begin{aligned} D_0 &= y_t \\ D_1 &= (y_t - y_{t-1}) \\ D_2 &= (y_{t-2}) - (y_t - y_{t-1}) \\ D_2 &= \underline{\underline{y_{t-2} - y_t + y_{t-1}}} \end{aligned}$$

Moving Average \Rightarrow Error

$$\hat{y}_t = \epsilon_{t-1} \psi_{t-1} + \epsilon_{t-2} \psi_{t-2} + \epsilon_{t-3} \psi_{t-3} + \dots + \epsilon_{t-n} \psi_{t-n} + C$$

$$\begin{array}{c} \text{Predicted} \\ \hat{y}_t \\ \text{Error} \\ | \hat{y}_t - y_t = \text{error} | \\ y_{t-1} \\ y_t \end{array}$$

$y_t \Rightarrow$ value at t time
or value q_t prekai
 $\psi \Rightarrow$ coefficient
 $C \Rightarrow$ constant
 $\epsilon \Rightarrow$ error

ARIMA \Rightarrow

$$\hat{y}_t = y_{t-1} \psi_{t-1} + y_{t-2} \psi_{t-2} + \dots + y_{t-n} \psi_{t-n} + C + (\epsilon_{t-1} \psi_{t-1}) +$$

PASF

$$\frac{\epsilon_{t-d} \psi_{t-d} + \dots + \epsilon_{t-n} \psi_{t-n} + C}{ACF}$$

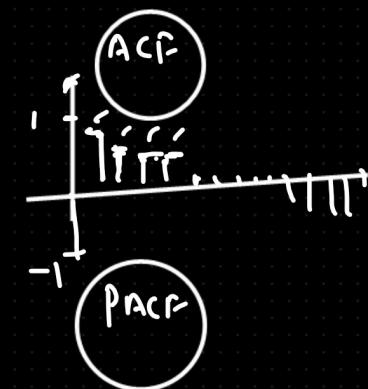
Python
(AR(1))
(AR(2))

D=0 ma(1)
D=1 ma(2)
D=2

$AR = p \Rightarrow \{0, 1, \dots, n\} \Rightarrow$
 $D \Rightarrow d \Rightarrow \{0, 1, \dots, n\} \Rightarrow$
 $m_A \Rightarrow q \Rightarrow \{0, 1, \dots, n\} =$

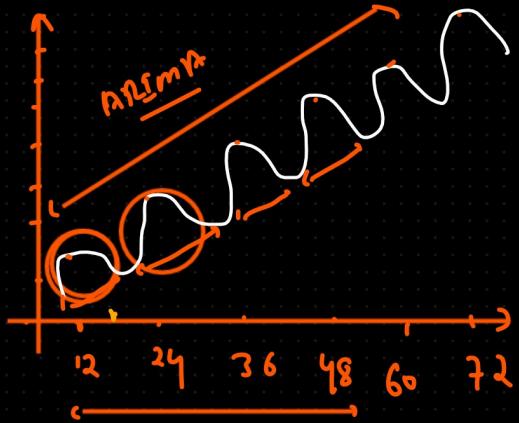
\downarrow
 $(ACF | PACF) \Rightarrow$ Corrlogram
 \uparrow
 \uparrow

4



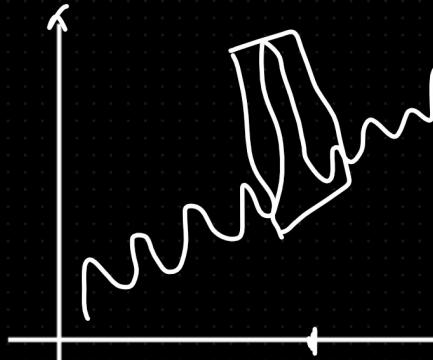
SARIMAX

Seasonal



SARIMAX

(outliers fraction)
exogenous



environl factor \Rightarrow Outliers

$AR \leq m_A$
 $\frac{1}{(P, d, q)}(P, D, Q) \times$

$(P, d, q) \times (P, D, Q)_S$

$AR = \{0, 1, 2, \dots, n\}$ $d = \{0, 1, 2, \dots, n\}$ $m_A \text{ and } q = \{0, 1, 2, \dots, n\}$



$$\rightarrow \frac{(P, d, q)}{\downarrow} \underset{\substack{nr \leq m \\ \downarrow}}{(P, D, Q)} \textcircled{2}$$

(wholedata) (seasonaldata)