

time series data.

diff ? —

t	10
t+1	20
t+2	30
t+3	40
t+4	50
	60
	70
	80
	90
	100

first diff
→ mean → not const
— std → not const

→ { 10
10
10
10
...
1 }

Machine learning

Linear regression.

Variable
 \Rightarrow

X Y
(height) (weight)

180	60
170	70
150	80
140	50
135	100

\Rightarrow

$$Y = mx + C$$

$$\text{Height} = m \times \text{Weight} + C$$

Simple linear reg.

$X_1 X_2 X_3 X_4 X_5 Y$

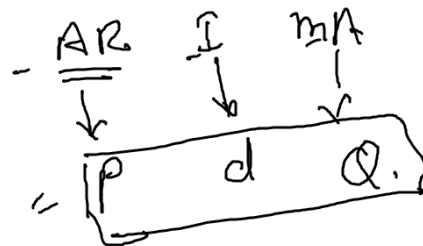
Lin. reg :-

$$Y = m_1 X_1 + m_2 X_2 + m_3 X_3 + m_4 X_4 + m_5 X_5 + C$$

multiple lin reg

(ARIMA)

Autoregression. (AR)



Previous time stamp

next time stamp regression with in a variable.

time

- 10(t-5)
- 20(t-4)
- 30(t-3)
- 40(t-2)
- 50(t-1)
- 60(t)
- 70(t+1)
- 80(t+2)
- 90(t+3)
- 100(t+4)

(1) train.

(2) test (validation)

X

X	Y
10	20
20	30
30	40
40	50
50	60
60	70
70	80
80	90
90	100

(1)

10
20
30
40
50
60

20
30
40
50
60
70

30
40
50
60
70

40
50
60
70
80

$$60 = 70$$

$$70 = 80$$

$$80 = 90$$

$$90 = 100$$

$$\text{ARIMA} :- Y_t = b + (t-1)m_1 + (t-2)m_2 + (t-3)m_3 + \dots + (t-k)m_k$$

multiple Unres.

Autorisier

$$Y_t = b + m_1(t-1) + m_2(t-2) + m_3(t-3) + m_4(t-4) + m_5(t-5)$$

= next
timestamp



Previous time
stamp

ARIMA

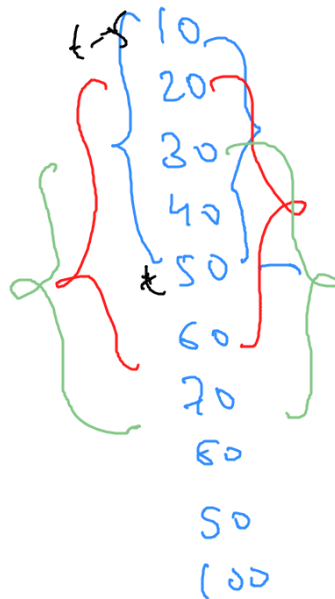


p d → difference

d = 1 One differencing.

d = 2 Second differencing.

5 rolling



NA
NA
NA
NA

30
40
50
⋮

MA → SMA, EMA, EWMA

formula

$$\frac{10+20+30+40+50}{5} = 30$$

$$\frac{20+30+40+50+60}{5} = 40$$

$$\frac{30+40+50+60+70}{5} = 50$$

$$y = (\epsilon_{t-1})$$

$$y = \theta + c(\epsilon_{t-1})$$

SARIMA

rolling window = $t-1$ to t

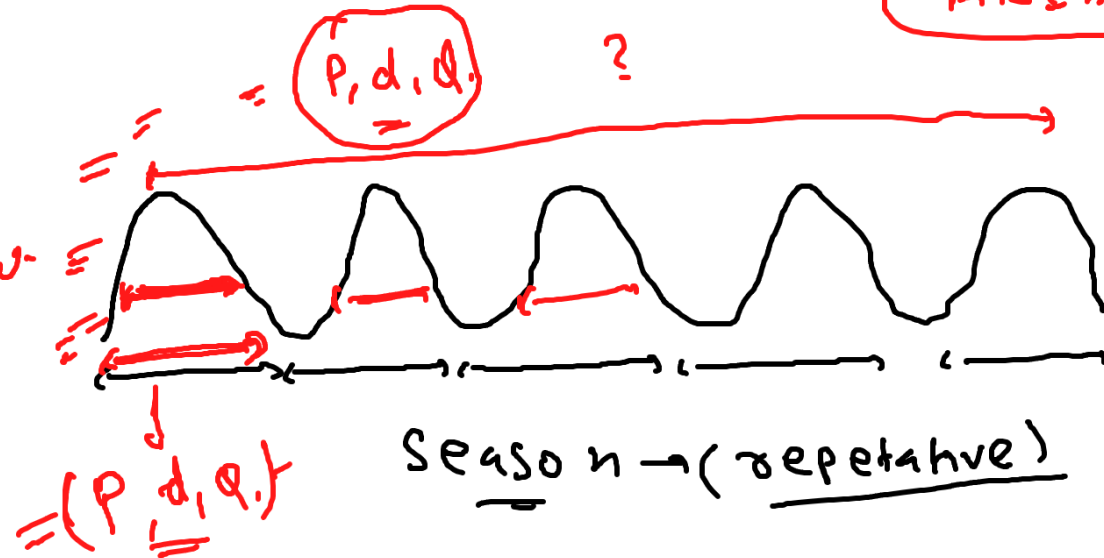
Seasonal ARIMA

Seasonal factor
+
ARIMA

AR :-

t
 $t-1$
 $t-2$
 $t-3$
 $t-4$

rolling window



SARIMA

ARIMA

SARIMA

SARIMAX

entire series

P, d, Q
=

$\begin{pmatrix} P=1 \\ d=2 \\ Q=1 \end{pmatrix}$

$\begin{pmatrix} S \\ P=0 \\ d=1 \\ Q=2 \end{pmatrix}$

X
12

(Count of season)
(frequency of season)

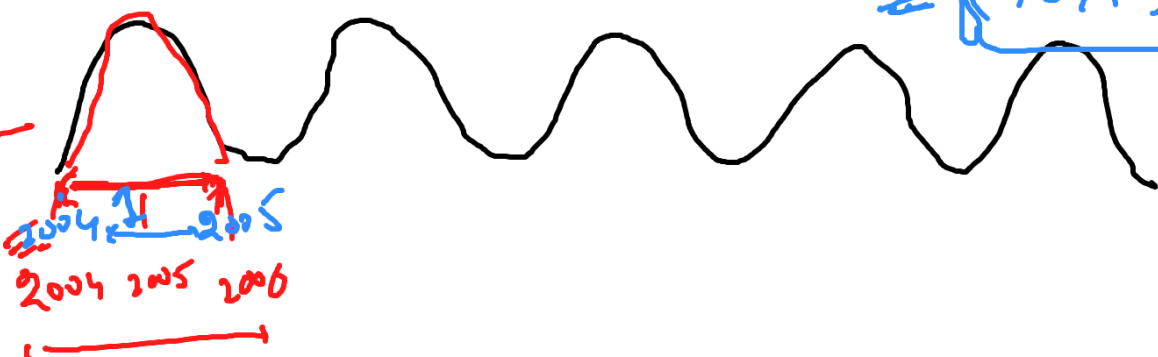
$(P, d, Q) [P, d, Q, X]$

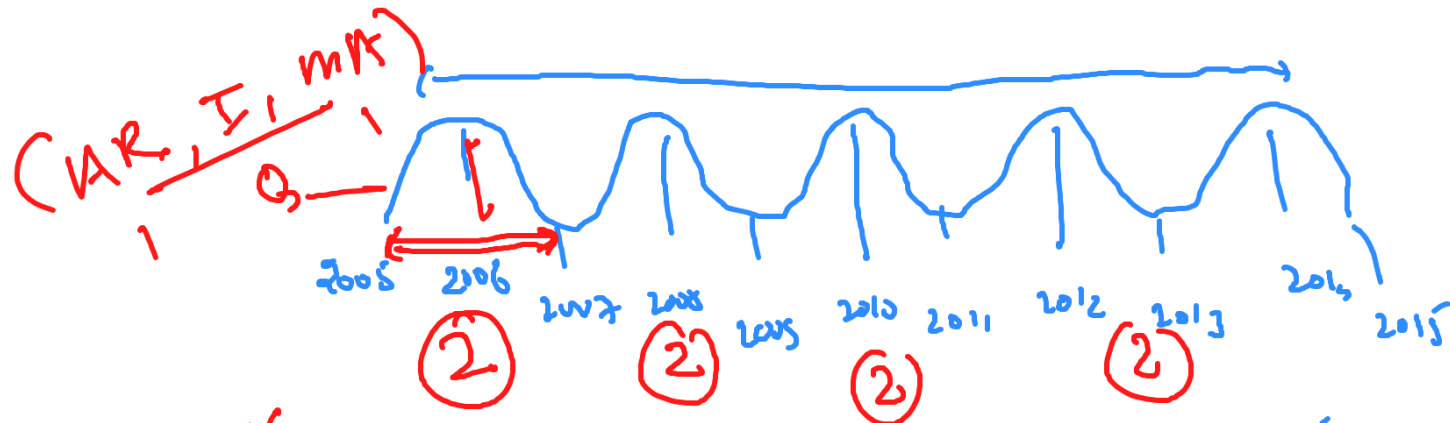
$\Rightarrow [(1, 2, 1) [0, 1, 2, 12]]$

month = 12, 18, 24, 36

P, d, Q

2 years





→ X
 Season factor in my model

10 year = $AR = P = 1$

$I = d = 0$ $(1, 0, 2) [1, 0, 1]$

$MA = Q = 2$

SARIMA

$(P, d, Q) [P, D, Q] \times$

$(1, 0, 2) [1, 0, 1] 2$

$(1, 0, 2) [1, 0, 1] (2)$

// Again, next, split
 0.25 0.30 0.50

h, p, q
 150 60 70 AR.

Previous time step
 1, 2, 3
 3, 4, 5
 4, 5, 6

$$P = AR.$$

rolling window

③ { 1
 2
 3
 4
 5
 6 }

ACF, PACF

t 10 20
 $t+1$ 20 30
 $t+2$ 30 40
 $t+3$ 40 50
 $t+4$ 50 60
 $t+5$ 60 70
 $t+6$ 70

Corr = 1

Auto [correlation] plot
 \downarrow \rightarrow \times $Y \rightarrow =$

10 20
 20 30
 30 40
 40 50
 50 60
 60 70
 70

$\rightarrow ?$

\times within a variable with respect to time lag Corr.

10:00 100
 11:00 20
 12:00 30
 13:00 40
 14:00 50

directly flat
 Partial.

