Association

↓

Transaction data → tranteaction database

↓

200 customers
X7 days
X 4 weeks
X 12 months
X 5 years

Transaction Database →

<

Supermarket, Kirana, Grocery, clothes ...

→ Association ⇒ Customer X product

Association → Combine A product

→ max sale

→ min sale

→ Combine sale

→ offer

1. Itemset → set of 1/more items

2. k - itemset → $\underline{X} = \{x_1, \ldots x_k\}$

↑
Today

customer → Transaction

3. Frequent pattern → what is a pattern ?

4. Association Rule → Association Rule Mining

↑

frequent itemsets

Use → Applications

1. Market – Basket data analysis
2. Cross – marketing
3. Offers
4. Sale – 40%

Association → correlation

→ patterns → Sequential

→ Structural

- Structural
  - Time Series
  - Spatiotemporal
- Classification → Frequent Pattern analysis
- Clustering → Frequent Pattern analysis

Eg

Data

Tid	Items
1.	A B C D
2	A C D
3	A B C

Transactions

Matrix Representation

	A	B	C	D	E
4 ← 1	1	1	1	1	0
3 ← 2	1	0	1	1	0
3 ← 3	1	1	1	0	0

Trans	
4	C D E
5	A B C E

3 ← 4	0	0	1	1	1
4 ← 5	1	1	1	0	1
	4	3	5	3	2

Iteration-1 $\epsilon = 2$

Candidate Item	Support
A	4
B	3
C	5
D	3
~~E~~	~~2~~

2 Comb

Iteration-2

Candidate	Support
AB	3
AC	4
~~AD~~	~~2~~
BC	3
~~BD~~	~~1~~
CD	3

3 Comb

Iteration-3

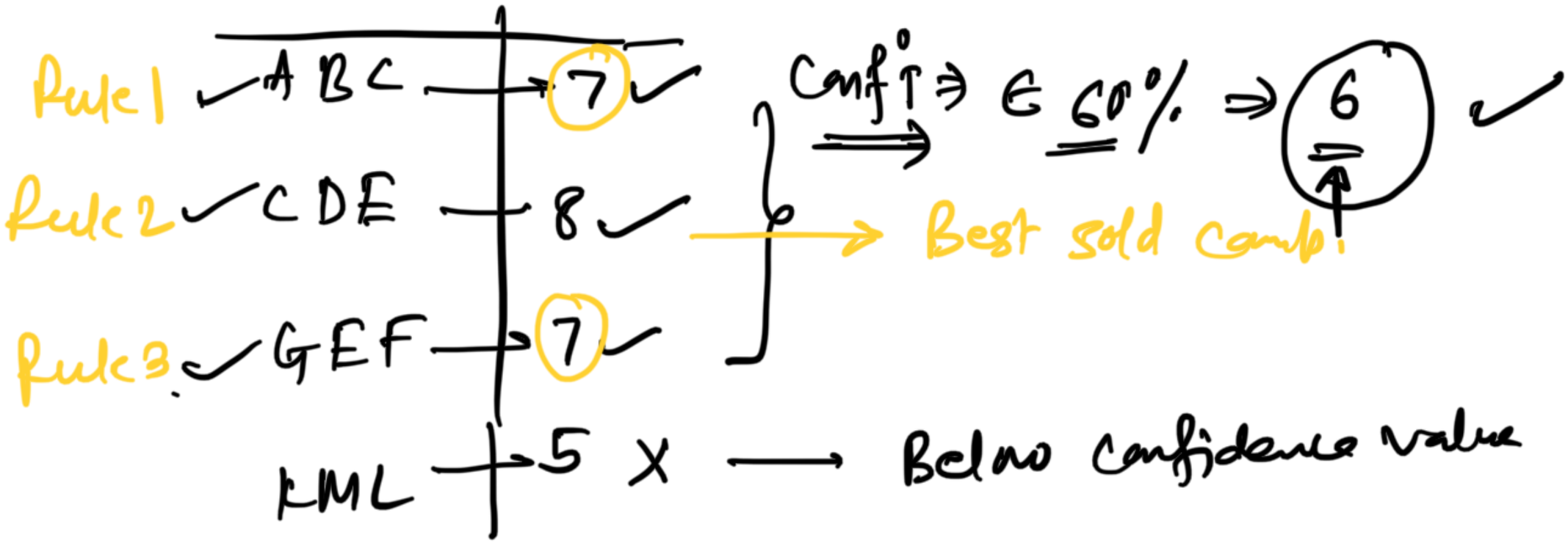
Candidate	Support
ABC	3
~~ABD~~	~~1~~
~~ACD~~	~~2~~

Comb ⇒ Associating 1 product with Another

~~ABC~~ → 3

Apriori Algorithm

Association Rule → Generate



Association Rule Mining

1. Bread, Milk
2. Bread, Diaper, Beer, Eggs
3. Milk, Diapes, Beer, Coke
4. Bread, Milk, Diapes, Beer

5. Coffee, Bread, Milk, Diaper, Coke

Association Rule

LHS → FP

RHS → Comb

{ Diaper } ⟶ { Beer } → Repeat

{ Milk, Bread } ⟶ { Diaper, Coke, Beer }

Count

{ Bread, Bees } → { Diapes, Milk, ~~Eggs~~ }

Frequent comb

Additional comb

Frequent Itemset

Itemset → collection of items (1/more items)

k-itemset → items in all / k transactions

Support $\epsilon = 2$ / $\sigma = 2$

- Fraction of transactions that contain an itemset

Support count = occurrence of an itemset

↓

Frequency of

$$\text{Support} = \frac{\text{freq}(X,Y)}{N}$$

Association Rule

$X \Rightarrow Y$

↓ → {k itemset}

→ o/p

$A \rightarrow B$

$P(A \cap B)$

{Frequent item}

$$\text{Confidence} = \frac{\text{freq}(X,Y)}{\text{freq}(X)}$$

$A \rightarrow B$

$P(B/A)$

$$\text{Lift} = \frac{\text{Support}}{\text{Support}(X)\ \text{Support}(Y)}$$

Ex. → $\{\text{Milk, Diaper}\} \Rightarrow \text{Beer}$

$X \Rightarrow Y$

$$\sigma\{\text{Milk, Diaper, Beer}\} \quad 2$$

$$S = \frac{\qquad\qquad}{5} = \frac{\qquad}{5}$$

$$C = \frac{\sigma(\text{Milk, Diaper, Beer})}{\sigma(\text{Milk, Diapers})} = \frac{2}{3} = 0.6$$

Time Series Analysis

— Sequence of dates $\begin{cases} \text{day} \\ \text{month} \\ \text{years} \end{cases}$

TS Analysis → analysis of data

↓

Come across units of time

DD:MM:YYYY:00:00:00 → formal for

TIME

↓

one variable

TIME SERIES

- day
- month
- year
- Morning -.-

→ order to extract meaningful information by using statistics

TSA

↑

important area in research

↙ univariate

↓

one-variable

↘ Multivariate

↓

many variables

→ App → Stock Market Analysis

- Medical Analysis
- Market Analysis
- Geographical study → rock, sea, volcanoes.

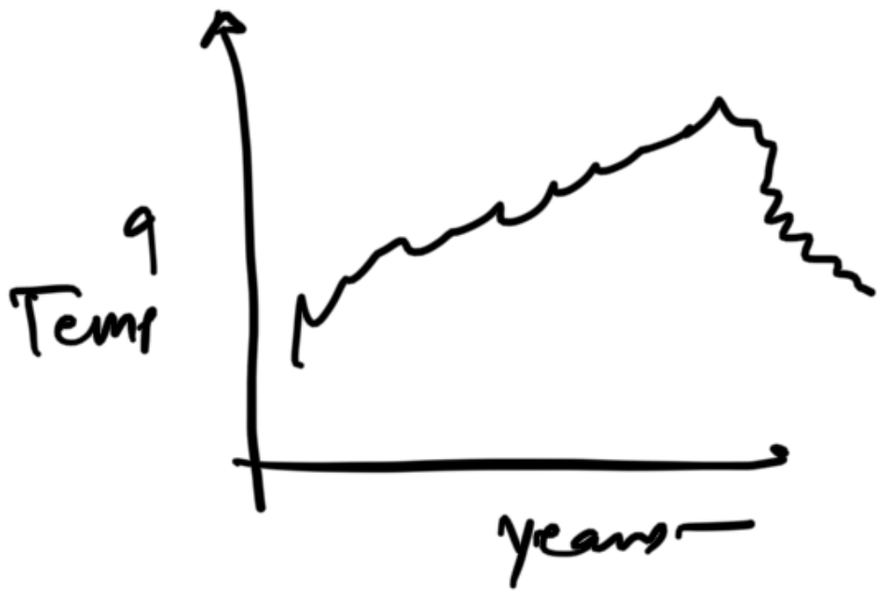
→ Time Serice Data

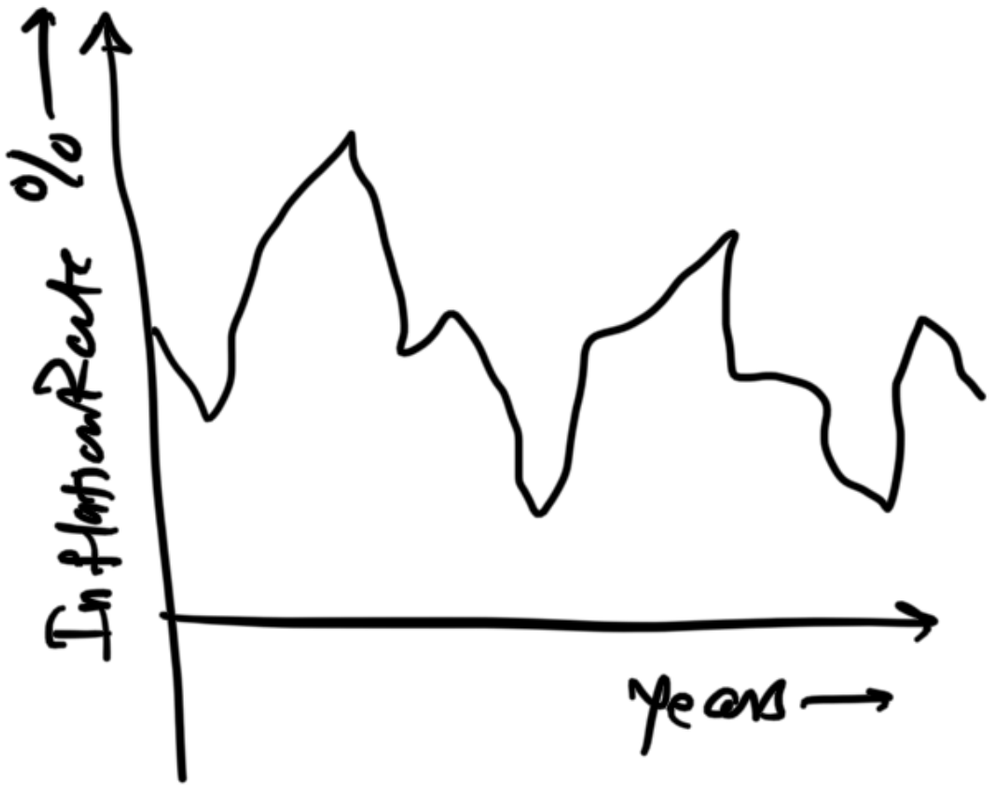
- → Numesic data
- → Regular time interval
- Intervals can be different formats
  - → annually
  - → quarterly
  - → monthly
  - → weekly

→ daily

→ hourly

Time Plot → 2D

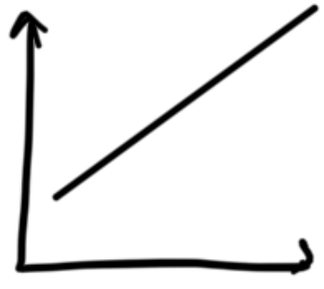


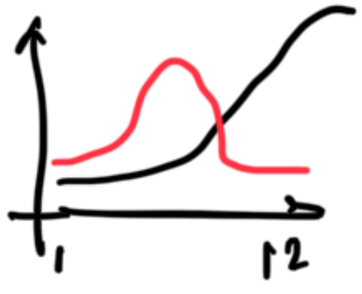


- equally spaced
- Cross sectional data → different for different individuals

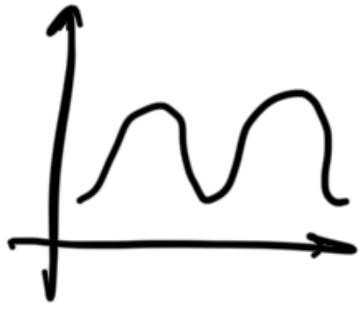
Time-Series Components →

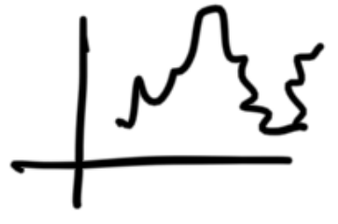
- Trend Component → Overall persistent
- Seasonal Component → Winter-4 months; Regular/Periodic intervals
- Cyclic Component → Repeating intervals
- Irregular Component → Erratic or random





— Winter





— Univ

Mathematically –

$$U_t = f(t)$$

U_t: value of variable under variation with time(t)

Time Series Analysis → Decomposition of Time Series

$$U_t = T_t + S_t + C_t + R_t$$

U_t = Time series value at time t

T_t = Trend

S_t = Seasonal

S_t = Seasonal

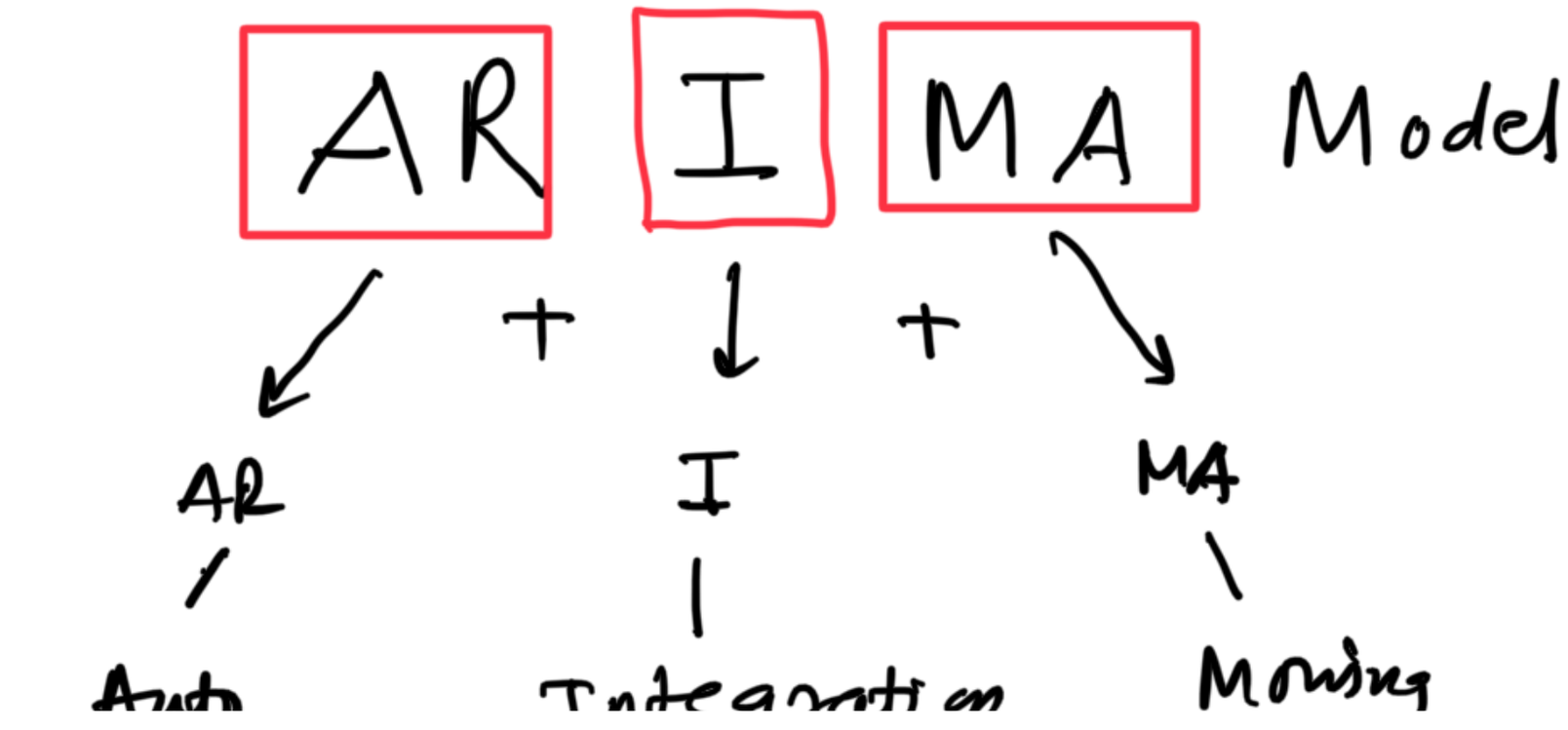
C_t = Cyclic

R = Random

Decomposition by Multiplication hypothesis

$$U_t = T_t * S_t * C_t * R_t$$

ARIMA Model (Stats)



Regressive

Average