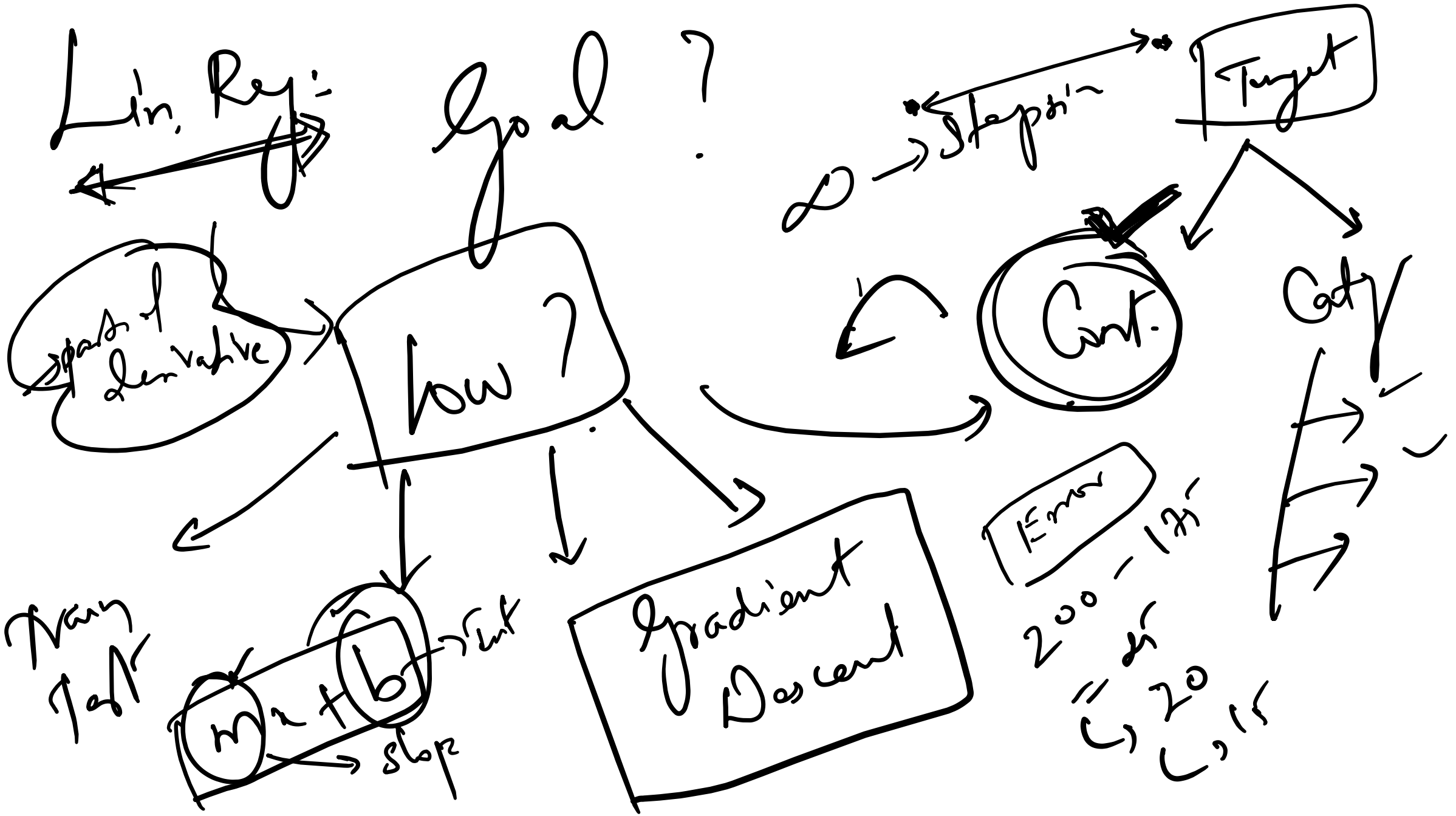
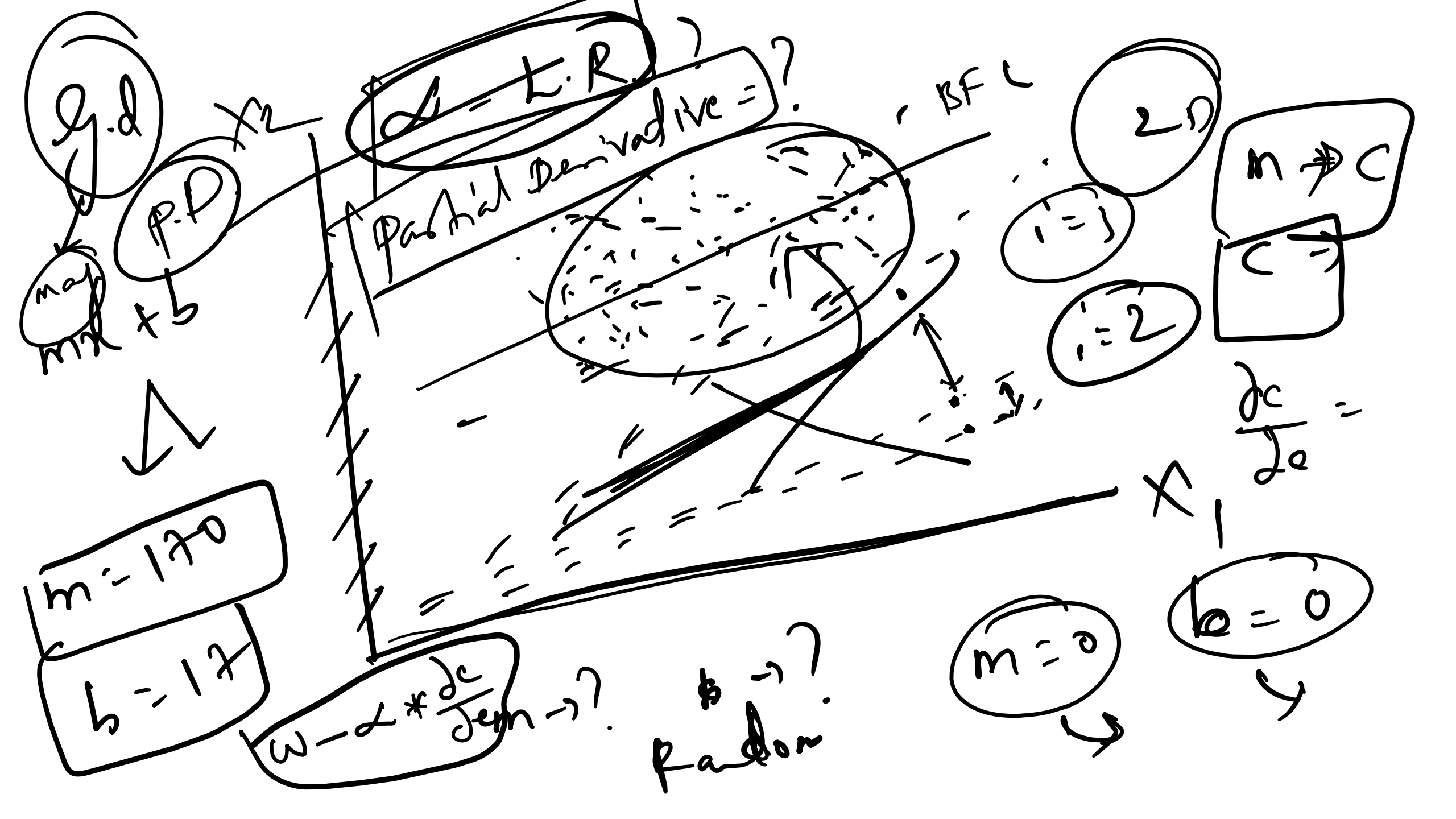
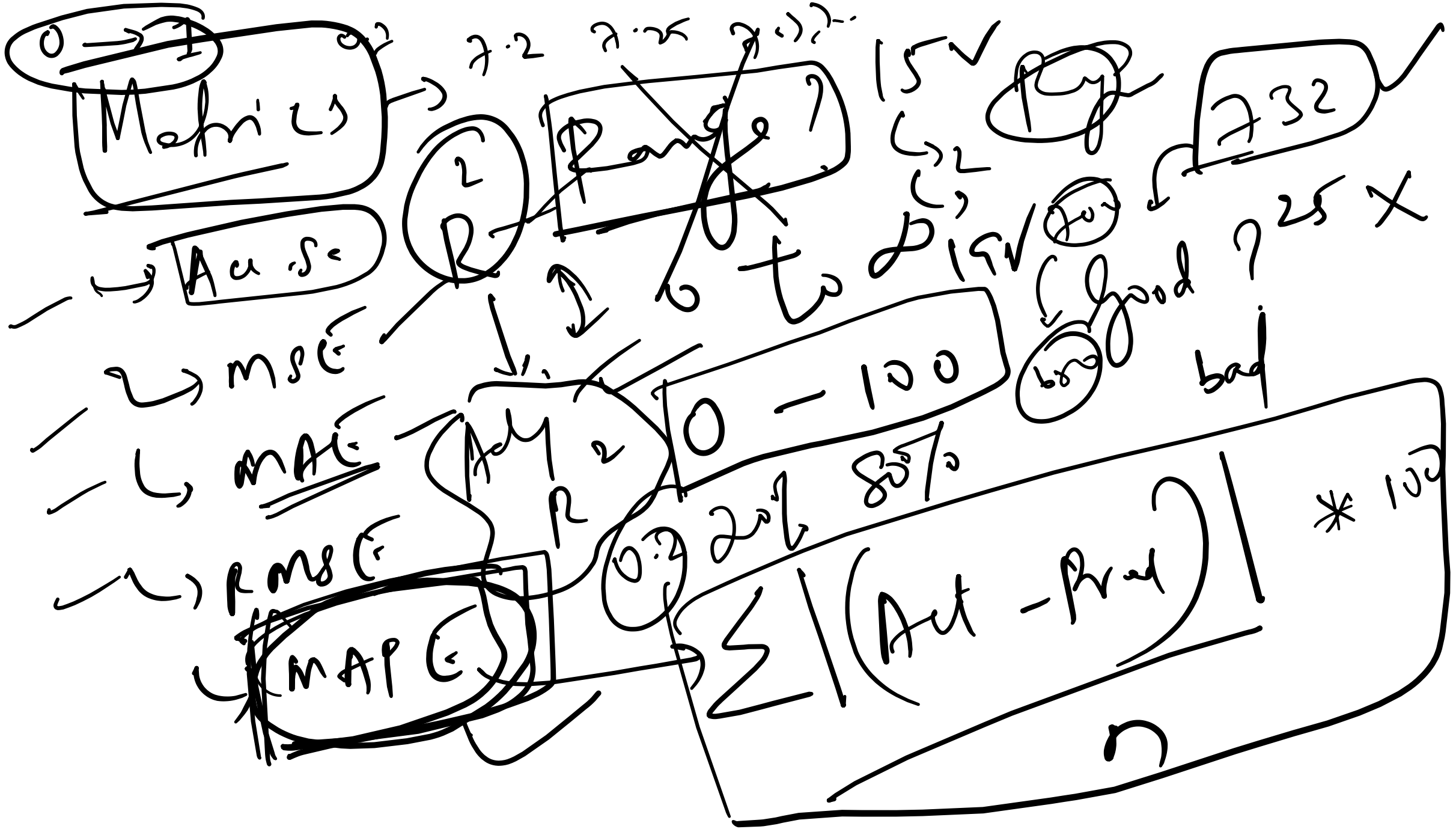
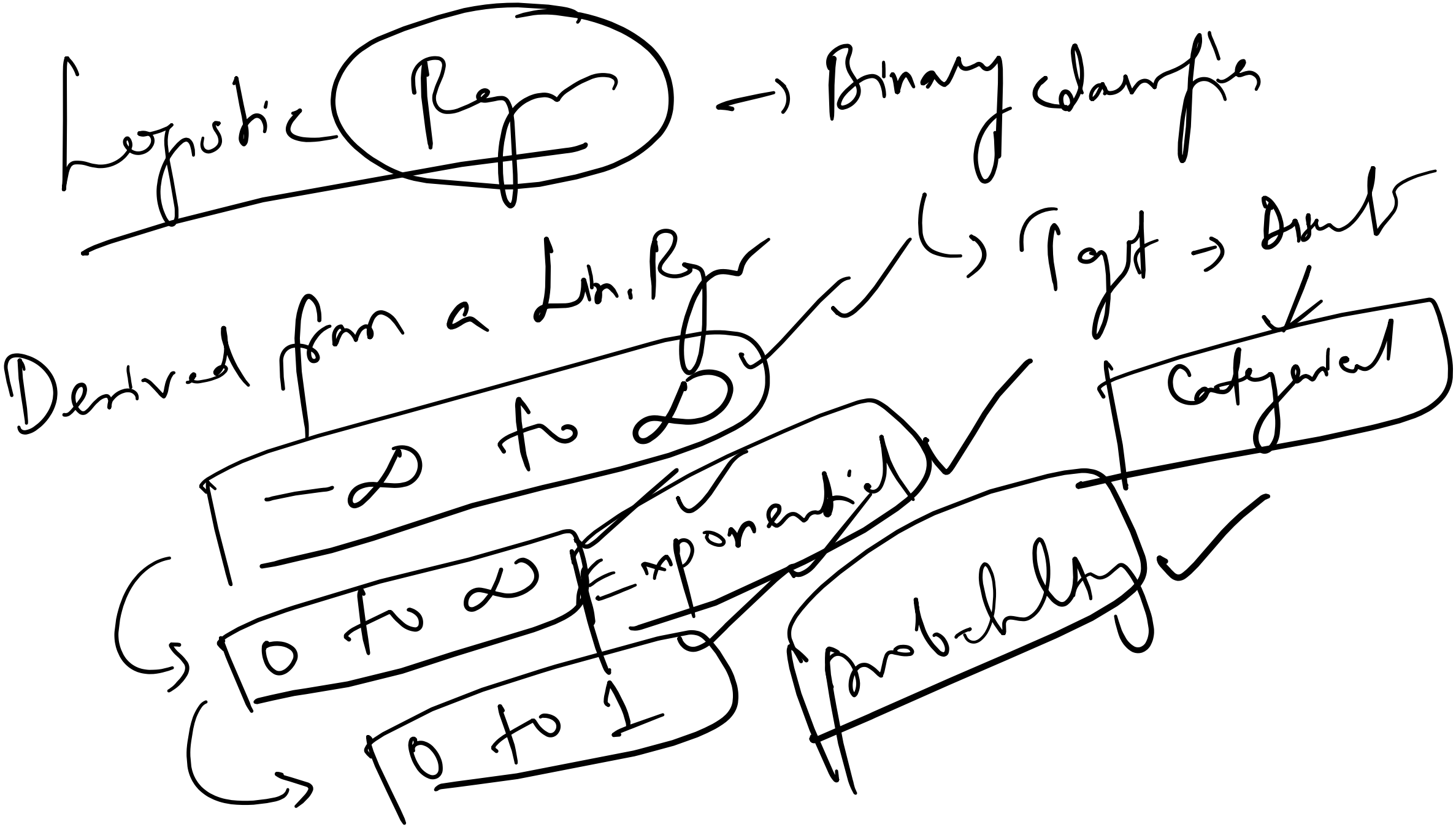
Algorithms ..

- Linear Regression — Regularization
- Logistic ..
- Ensemble
  - D-T
  - R.F
- N.B.
- KNN
- Boosting — Voting — Stacking


Lin. Reg :-
Goal ?
Steps
Target
Partial Derivative
How ?
Cost
Cost
Error
Gradient Descent
m x + b
slope


G.d
P.D
L = L.R
Partial Derivative = ?
BFL
2D
m ≫ C
C
i = 1
i = 2
mx + b
m = 170
b = 17
w − α * dc/dw → ?
b → ?
Random
m = 0
b = 0
dc/db =
X1


0 → 1
Metrics
Accuracy
MSE
MAE
RMSE
MAPE
R2
Range ?
0 - 100
0.2 or 80%
Adj R2
15 ✓
15 ✓
15 ✓
732
Good ?
bad
25 X
Σ |(Act - Pred)| * 100
n


Logistic Reg
→ Binary classification
Derived from a Lin. Reg
→ Tgt ⇒ Discrete
−∞ to ∞
0 to ∞
Exponential
Categorical
0 to 1
Probability

D.T → Overfitting ✗

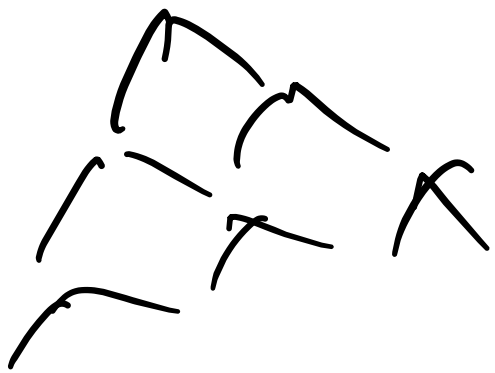
↓

R.F

① Random Variable

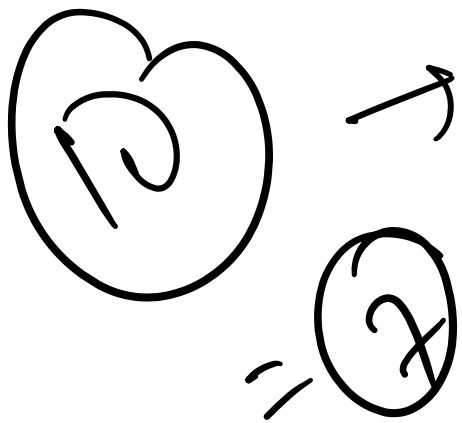
② Random subset of sample

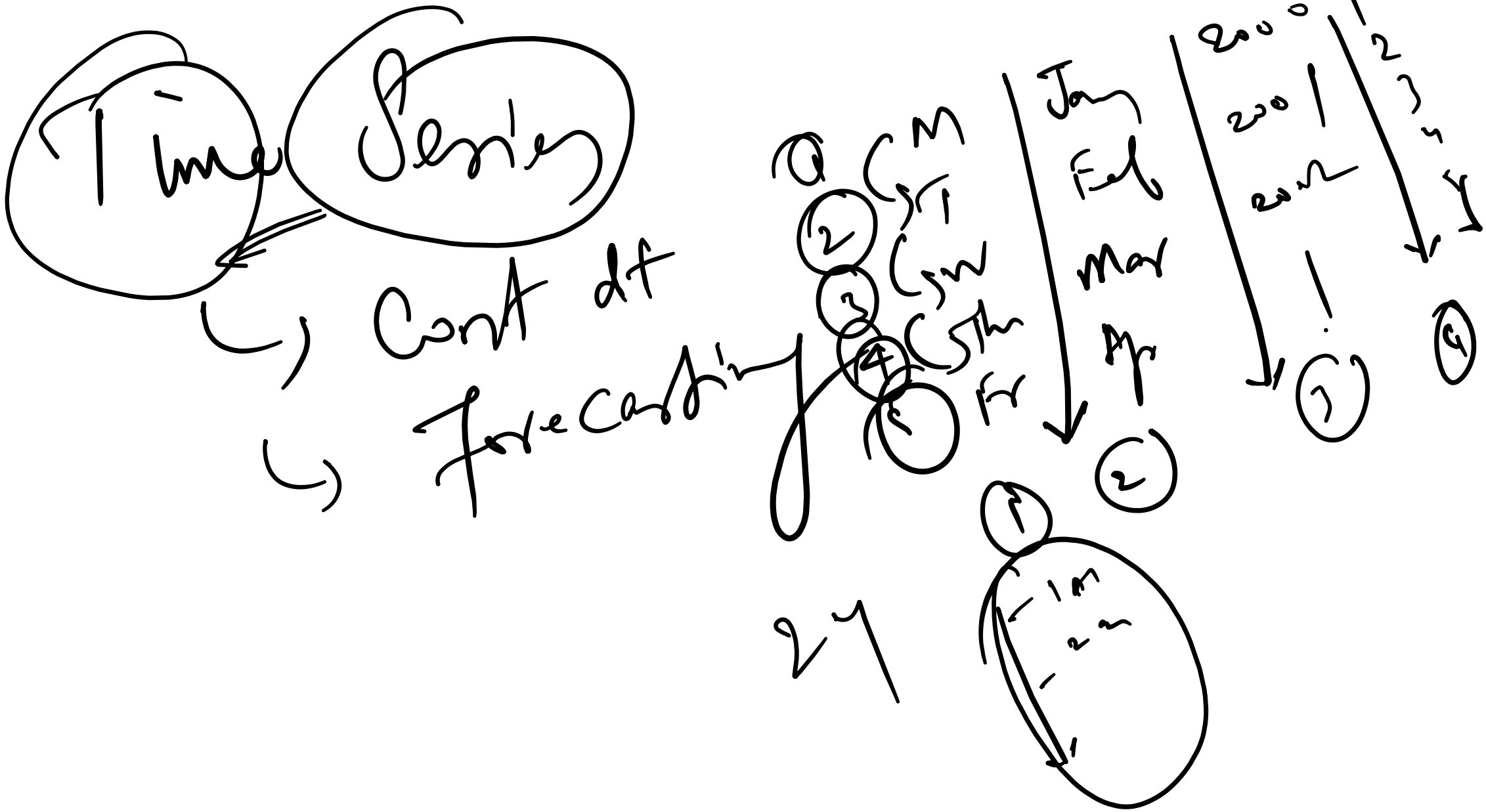
= -1



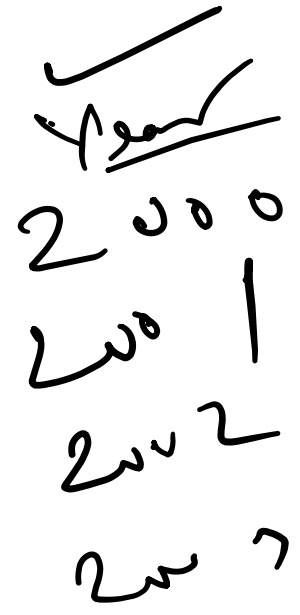
n_est :-

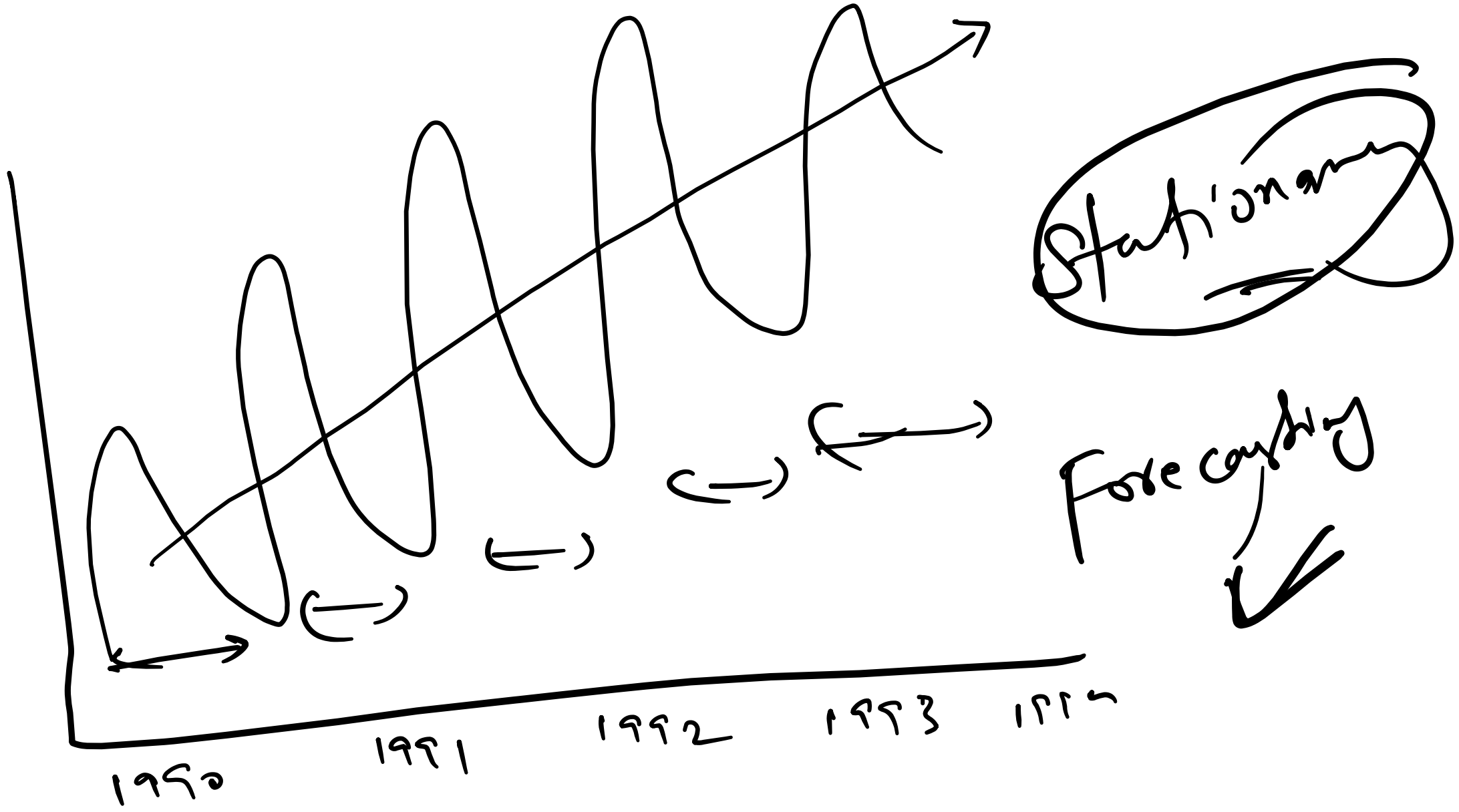
0.7 max_features

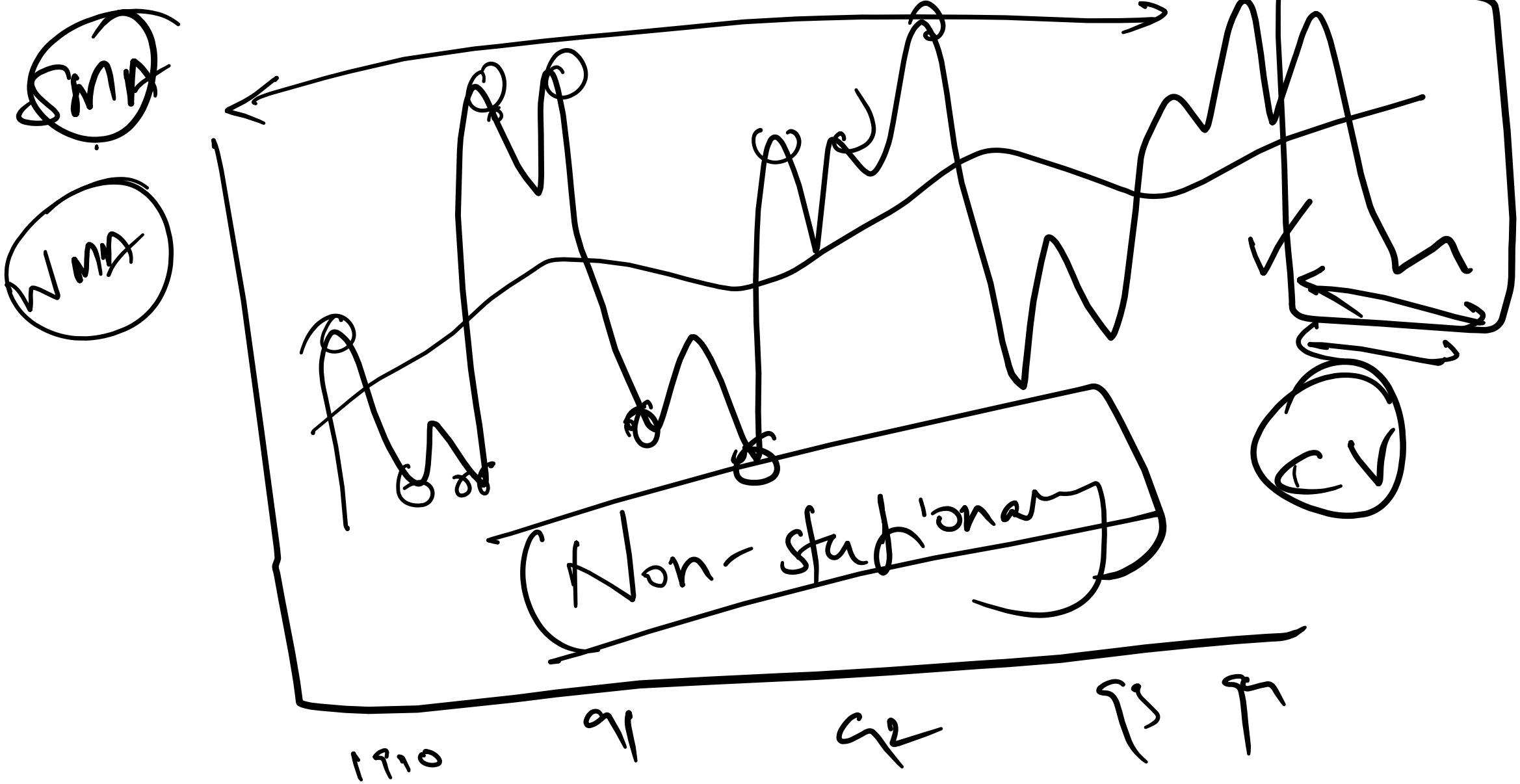



Time Series
Cont df
Forecasting
Jan
Feb
Mar
Apr
2000
2001
2002
M
T
W
Th
Fr
27
1M
2M

1. TREND
2. SEASONALITY
3. RANDOMNESS (RESIDUE / ERROR)




Stationary
Forecasting
1990
1991
1992
1993


SMA
WMA
Non-stationary
CV
1910
Q1
Q2
Q3
Q4

1996 → 7

2000 → 5

2001 → 4

2002 → 6

2003 → 5

2004 → 7

3 2 1

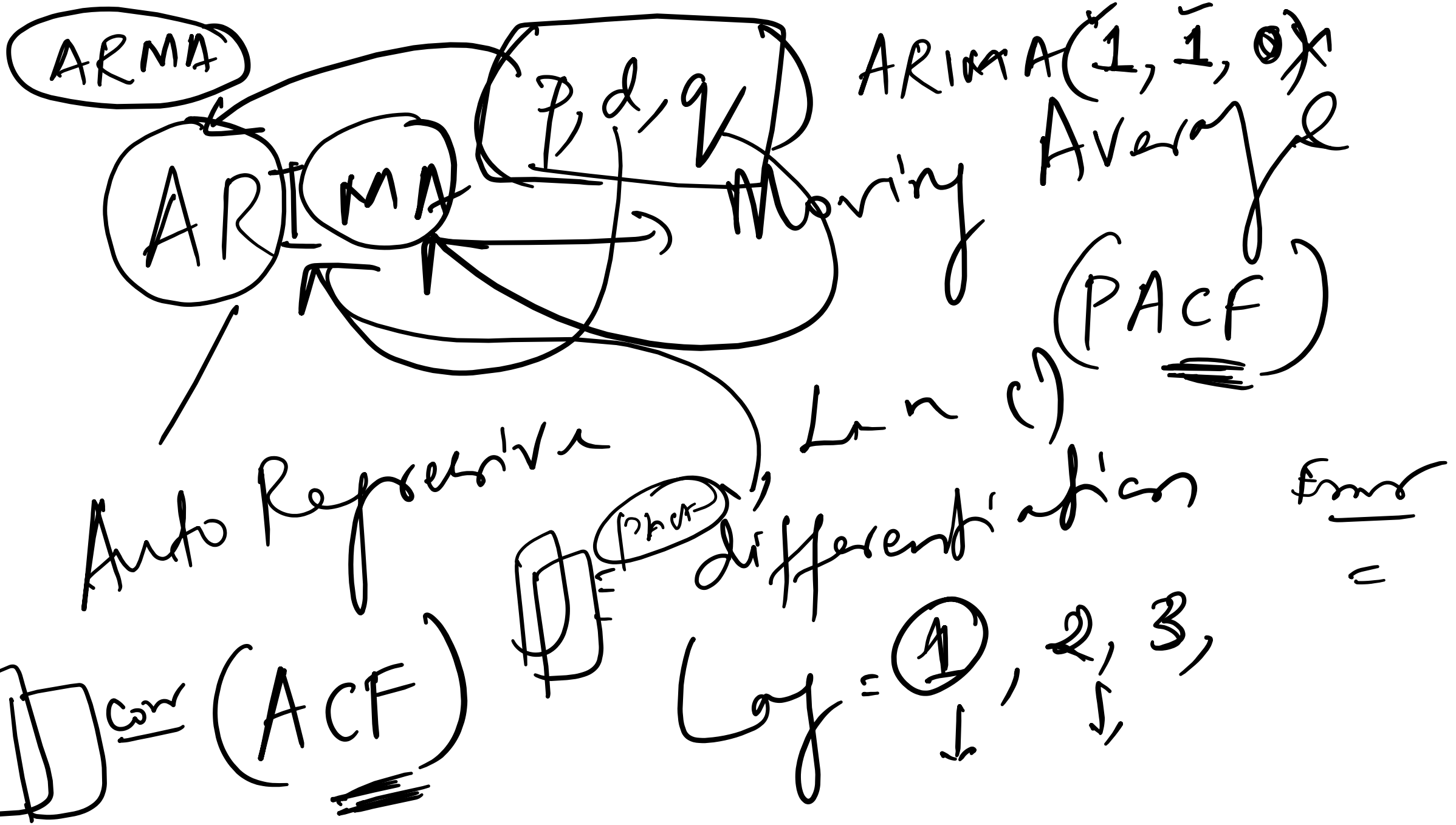
WMA(3)

WMA(4)

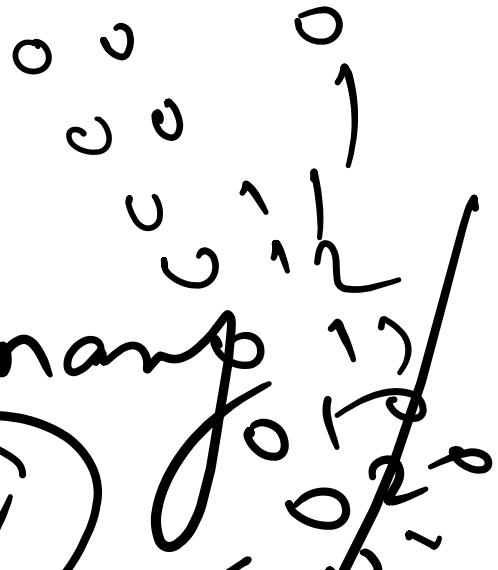
SMA(3)

$$\frac{5+4+6}{3}$$

$$\frac{6*4+4*3+5*2+7*1}{10}$$


ARMA
ARIMA
P, d, Q
Moving
Average
Auto Regressive
Integrated
(PACF)
(ACF)
Lag = 1, 2, 3,

Auto-Arima



I → Stationary / Non-Stationary

best p, d, q

p, d, q

p, d, q ⟹ ACF, I, PACF =

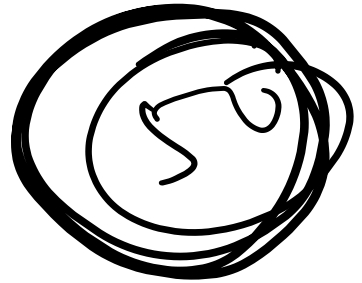
ACF: 1, ✓2, 3,

I: 0, 1 Non, 2,

PACF: 1, 2, 3

→ Grid Search CV

→ Randomized Search CV *



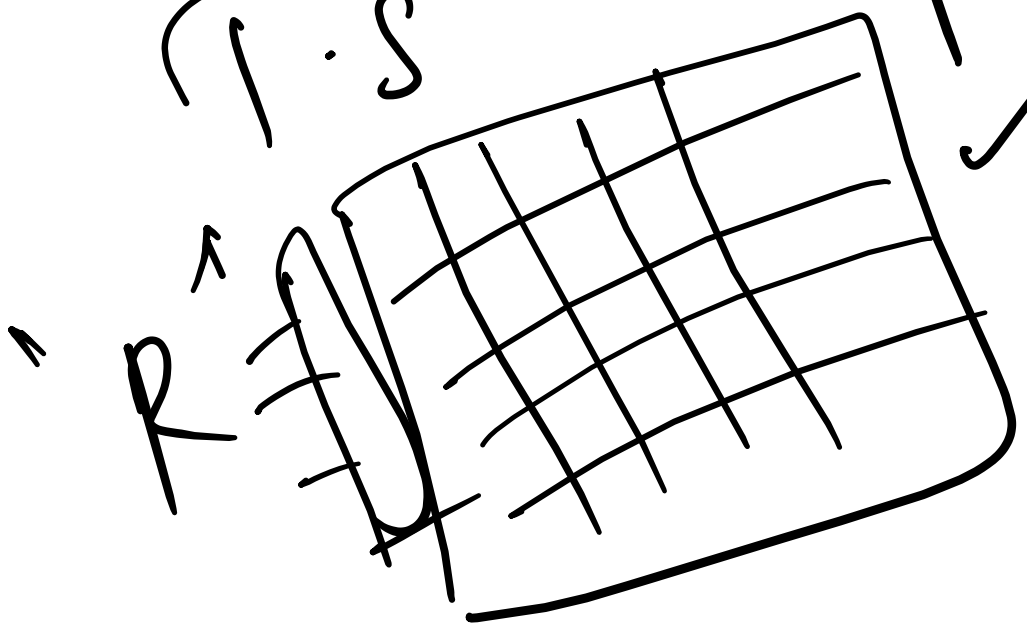
① n_est = [100, 200, 300, 400]

② max_feat = [0.3, 0.5, 0.6, 0.7]

Approach -

DECOMPOSITION

T.S → pre-processing



ARIMA

Holt winters

Exponential