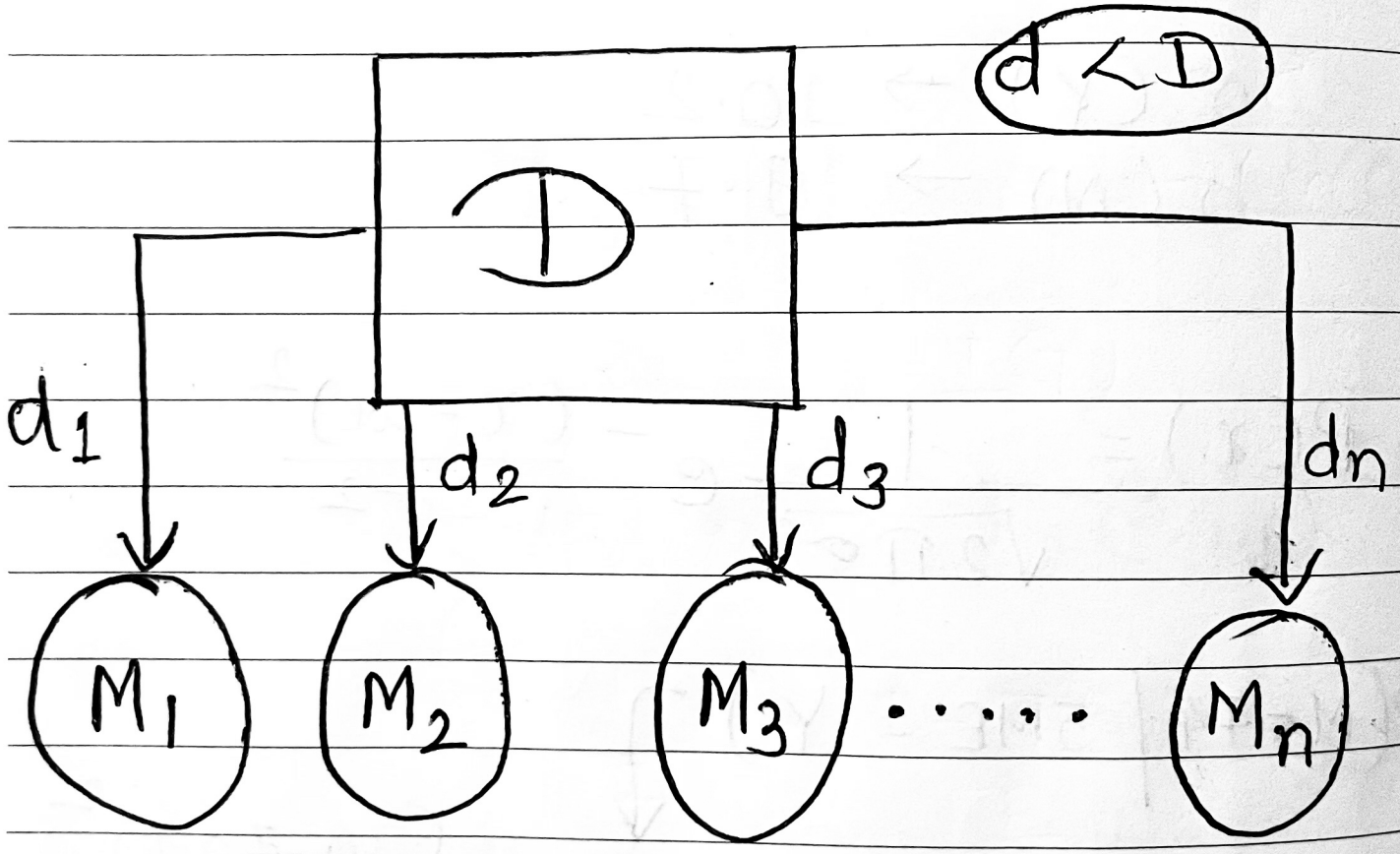
Ensemble Techniques

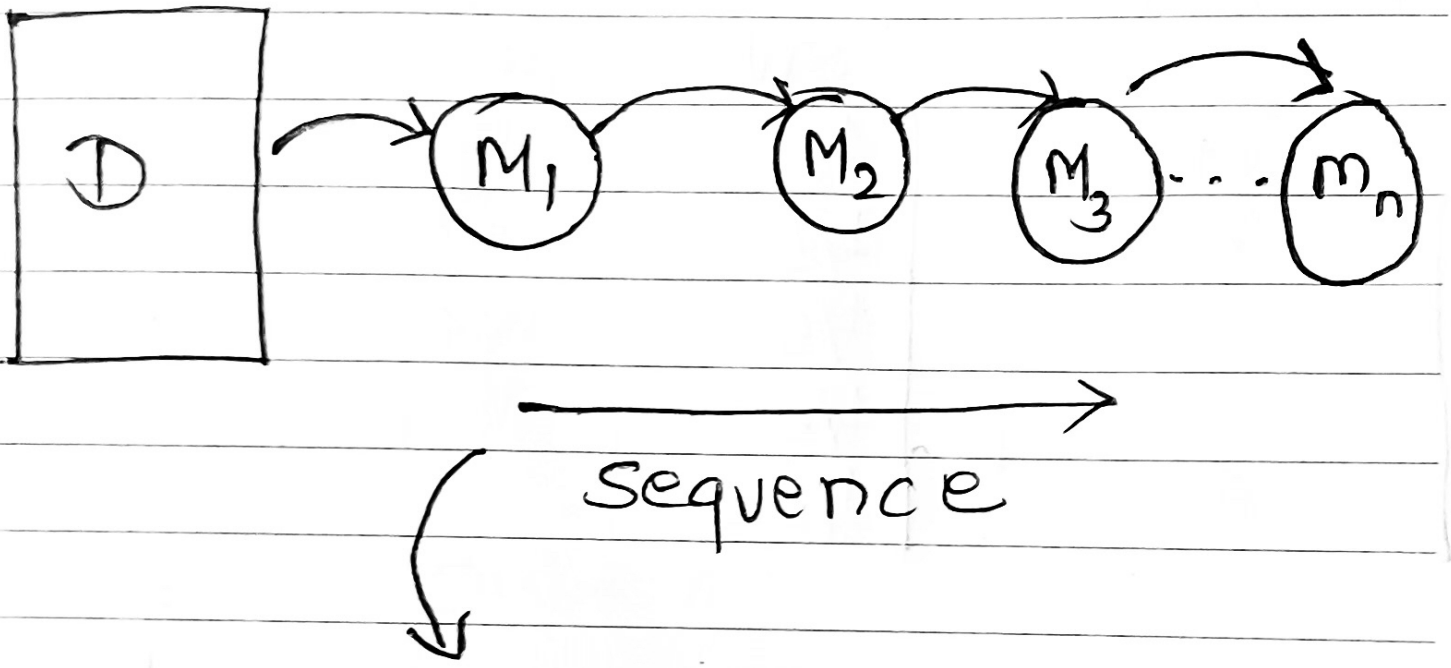
- Bagging → RF
- Boosting → AB
  - GB



Classification ⟶ Majority

Regression ⟶ Average

Boosting



$$\alpha_1 \times M_1 + \alpha_2 \times M_2 + \alpha_3 M_3 - - - + \alpha_n M_n$$

$\alpha \rightarrow$ weights / Importance / Influence

① Dataset

	"SW"
1	1/5
2	1/5
3	1/5
4	1/5
5	1/5

② Classify & get the Best (m=3) ← features

→ DT (lowest IG)

③ Calculate the 'α'

$$\frac{1}{2} \log \frac{1 - TE}{TE}$$

TE → Miss-classified

$$\text{Performance} = \frac{1}{2} \log \frac{1 - TE}{TE}$$

$$\alpha = \frac{1}{2} \log \frac{1 - 1/5}{1/5}$$

$$\boxed{\alpha = 0.69}$$

(4) New SW

		SW	NSW	NNSW
①		1/5	0.10	0.126
②		1/5	0.10	0.126
③	w	1/5	0.39	0.493
④		1/5	0.10	0.126
⑤		1/5	0.10	0.126
			0.79	0.997

$$NSW = OW * e^{\pm \alpha}$$

$- \alpha \longrightarrow$ Correct

$+ \longrightarrow$ Wrong

$$NSW_{(correct)} = \frac{1}{5} * e^{-0.69}$$

$$= 0.1004$$

$$NSW_{(wrong)} = \frac{1}{5} * e^{+0.69}$$

$$= 0.3988$$

(5) Bins / Bucket

	NNSW	Bins
①	0.126	0 - 0.126
②	0.126	0.126 - 0.252
③	0.493	0.252 - 0.745
④	0.126	0.745 - 0.871
⑤	0.126	0.871 - 0.997

(6) Random selection (0 - 1)

5

→ 0.011 → ①

→ 0.33 → ③

→ 0.45 → ③

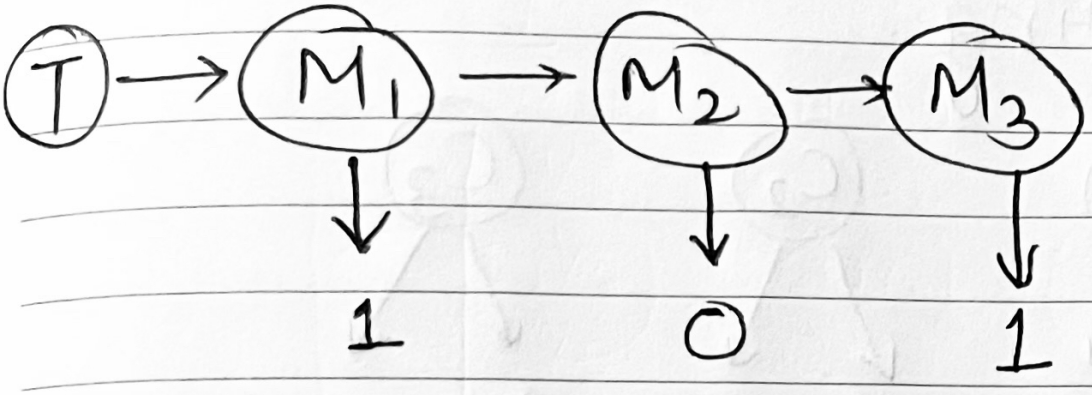
→ 0.60 → ③

→ 0.90 → ⑤

3 Times

①
③
⑤

New dataset



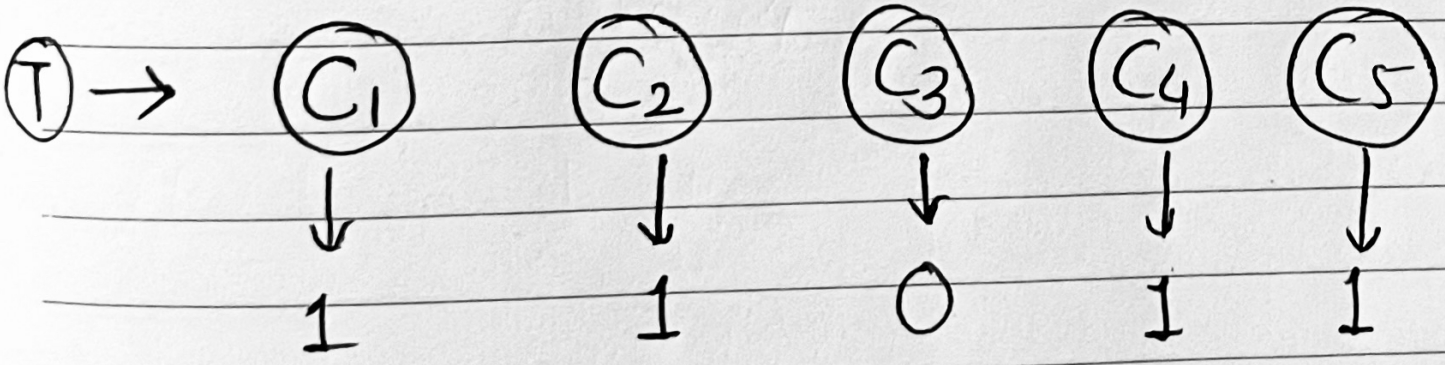
[redacted] → 0.69(1) + 0.51(0) + 0.30(1)

→ 0.99(1) + 0.51(0)

POS(1) → 0.99 ✓

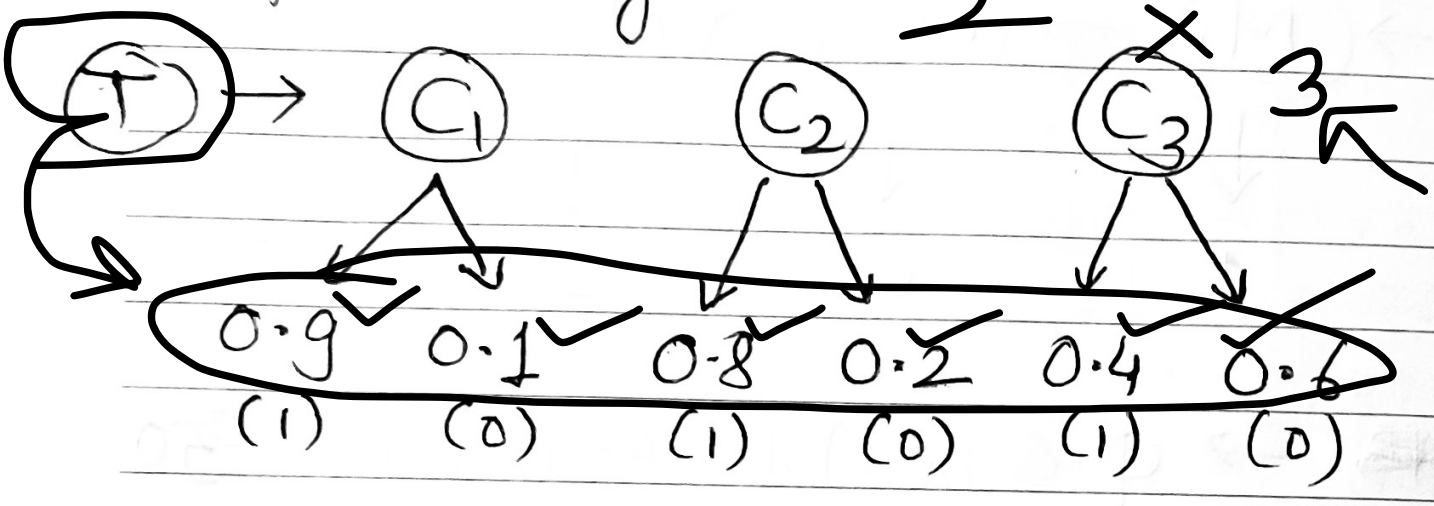
POS(0) → 0.51

Hard voting



Prediction → '1' (Majority)

Soft voting. → 2 1 0 / 3



(1) → $\frac{0.9 + 0.8 + 0.4}{3} = 0.7$ ✓

(0) → $\frac{0.1 + 0.2 + 0.6}{3} = 0.3$

Prediction → 1