Boosting + Usecase

19 December 2022 16:45

[0.6 - 0.8] .

XGBoost → Column Sampling, Row Sampling. Col Sample by tree, Col sample by level, Col sample by node

→ Parallelization.

→ Regularization → (reg_alpha), reg_lamda

0.1 — L2 regul — 0.1 — L1 Regularization

→ optimization

→ n_estimators

→ Learning rate (eta)

(light GBM)

Boosting

$\alpha_0 m_0 + \alpha_1 m_1 + \alpha_2 m_2 + \dots$ (Base learner)

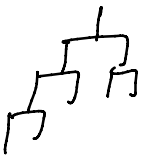
m_0 ← 0.6

(high Bias)

off from reality

{Shallow tree} ← (max_depth : 12)

299



System Resources — Train 100, learning 0.1 — (online)

Cv. = 5

Parameters (20) (20 x 5) (10) ←

Boosting

M → m_0

M → $\alpha_0 m_0 + \alpha_1 m_1$

M → $\alpha_0 m_0 + \alpha_1 m_1 + \alpha_2 m_2$

M → $\alpha_0 m_0 + \alpha_1 m_1 + \alpha_2 m_2 + \alpha_3 m_3$

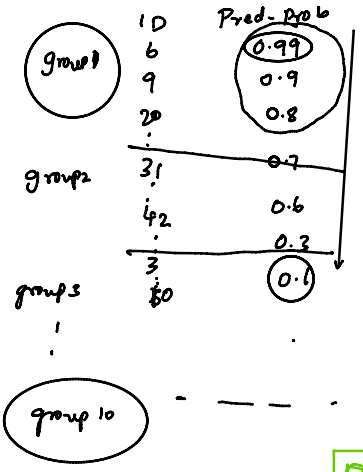
(n tree : 60)

60 → (M) → $\alpha_0 m_0 + \alpha_1 m_1 - - - \alpha_{60} m_{60}$ we this one

100 → (M) → $\alpha_0 m_0 + \alpha_1 m_1 + - - - - \alpha_m m_{100}$ (overfit) X] X

"Probability of Conversion"

Probabilities :-



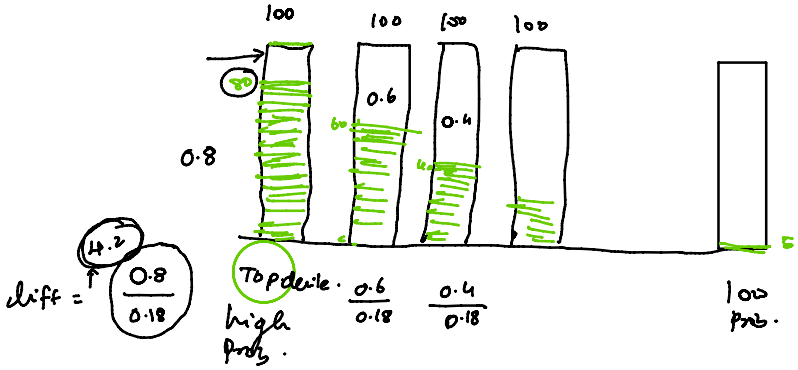
Sort descending (Predicted prob)

n = 1000

decile = 100

divide them into ten equal groups (deciles) ←

Decile chart



max 82% | minimum 18

Lift, Gain

P(conversion) ? = 0.18

if Randomly pick 100 quotes, how many of them do you expect to Convert ?

~ '18'

n = 100/10 = 10

	Ceil	Group
1/10	0.1	1
2/10	0.2	1
3/10	0.3	1
4		1
⋮		1
10/10	1	1
11/10	1.1	2
12/10	1.2	2
⋮		2

$\frac{11}{10}$ 1.1 2

$\frac{12}{10}$ 1.2 2

1 2

2

20 2 2

---

21 2.1 3

1 1 3

⋮ ⋮ 3

30 3 3

1

91

100