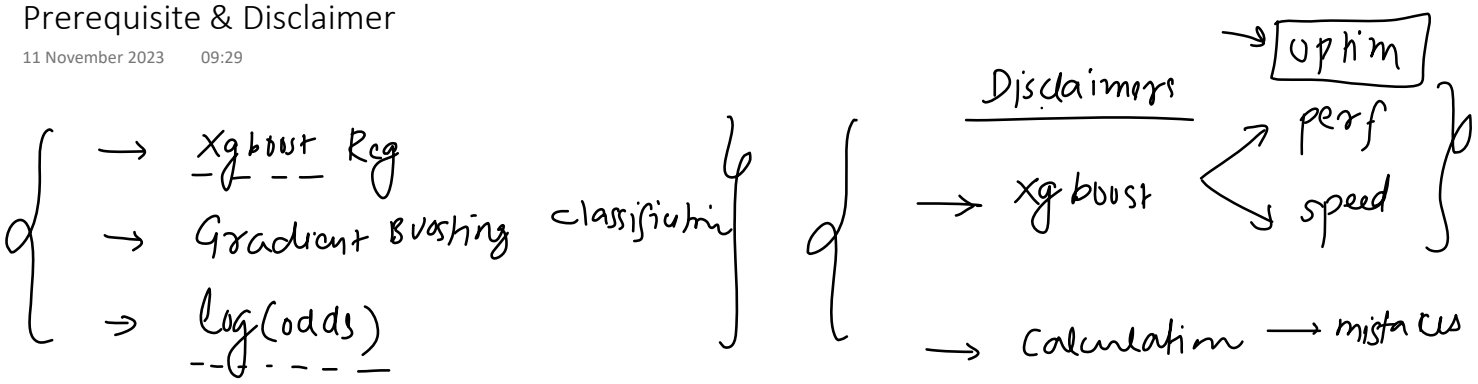
Prerequisite & Disclaimer

11 November 2023 09:29



Problem Statement

11 November 2023 09:29

cgpa	placed?
5.70	0
6.25	1
7.10	0
8.15	1
9.60	1

xgboost → classification

6.61 → 1,0

→ overview

→ step by step

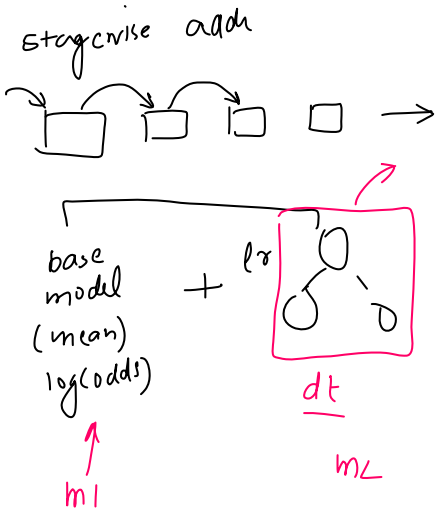
Overview

11 November 2023 09:31

cgpa	placed?
5.70	0
6.25	1
7.10	0
8.15	1
9.60	1

xgboost → GBDT →

Stagewise add



m1 m2 m3 . . . mn

GBDT → Vanilla dt { entropy / gini }

xgboost → diff dt { similarity score }

↳ xgboost reg

Step by Step Solution

11 November 2023 17:30

Stage 1 — □ → log(odds) → prob

cgpa	placed?	pred1(lo)
5.70	0	0.405 ✓
6.25	1	0.405
7.10	0	0.405
8.15	1	0.405
9.60	1	0.405

base estimator → mean

$\log(\text{odds}) \rightarrow \log\left(\frac{p}{1-p}\right)$ $p \rightarrow$ prob +ve class (1)

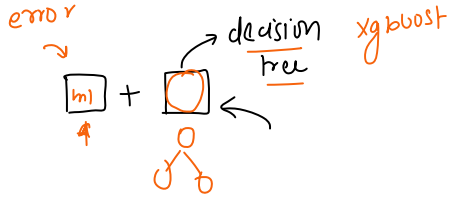
$\log\left(\frac{3/5}{2/5}\right) = \log_e\left(\frac{3}{2}\right) = 0.405$

$p = \frac{e^{\log odds}}{1+e^{\log odds}} = \frac{e^{0.405}}{1+e^{0.405}} = 0.60$

cgpa	placed?	pred1(lo)	pred1(prob)	res1
5.70	0	0.405 ✓	0.6	−0.6
6.25	1	0.405	0.6	0.4
7.10	0	0.405	0.6	−0.6
8.15	1	0.405	0.6	0.4
9.60	1	0.405	0.6	0.4

stage 1 → errors

pseudo residual



split → splitting criteria ←

	cgpa	res1
① 5.97	5.70	−0.6 ←
	6.25	0.4
② 6.67	7.10	−0.6
③ 7.62	8.15	0.4
④ 8.87	9.60	0.4

leaf node → [−0.6, 0.4, −0.6, 0.4, 0.4] → similarity score

SS = 0 ← maximize

$SS = \left(\sum \text{residual}_i\right)^2 \rightarrow$

GBDT Classification

$\frac{\sum \text{prev\_prob}_i (1-\text{prev\_prob}_i) + \lambda}{}$ → reg λ = 0

$= \frac{(-0.6+0.4-0.6+0.4+0.4)^2}{5 \times [0.6(1-0.6)] + 0} = 0$

gain → 1.67

	cgpa	res1
5.97	5.70	−0.6 ←
	6.25	0.4
6.67	7.10	−0.6
7.62	8.15	0.4
8.87	9.60	0.4

Splitting criteria = 5.97

cgpa < 5.97 ✓

−0.6	0.4, −0.6, 0.4, 0.4
SS_L	SS_R
$\frac{(-0.6)^2}{0.6(1-0.6)}$	$\frac{(0.4-0.6+0.4+0.4)^2}{4 \times 0.6 \times 0.4}$

Splitting criteria = 6.67

cgpa < 6.67 ✗

−0.6, 0.4	−0.6, 0.4, 0.4
SS_L	
$\frac{(-0.2)^2}{2[0.6 \times 0.4]}$	$\frac{(0.2)^2}{3 \times 0.24}$

8.15 | 0.4

8.87 { 9.60 | 0.4

$\frac{(-0.6)^2}{0.6(1-0.6)} = \frac{0.36}{0.24} = 1.5$

$\frac{0.36}{4 \times 0.6 \times 0.4} = \frac{1.5}{4} = 0.37$

$\frac{}{2[0.6 \times 0.4]} = \frac{0.04}{2 \times 0.24} = \frac{0.04}{0.48} = 0.08$

$\frac{}{3 \times 0.24} = \frac{0.04}{3 \times 0.24} = 0.05$

gain = $(SS_L + SS_R) - SS_{root}$

$= (1.5 + 0.37) - 0$

$= 1.87$

gain = $0.08 + 0.05 - 0$

$= 0.13$

cgpa	res1
5.70	−0.6 ←
6.25	0.4
7.10	−0.6
→ 7.62	0.4
8.15	0.4
8.87	0.4
9.60	

5.97 = (5.70 + 6.25)/2, 6.67 = (6.25 + 7.10)/2, 7.62 = (7.10 + 8.15)/2, 8.87 = (8.15 + 9.60)/2

Splitting criteria = 7.62

cgpa < 7.62

−0.6, 0.4, −0.6 | 0.4, 0.4

$\frac{(-0.8)^2}{3 \times 0.24} = 0.88$

$\frac{(0.8)^2}{2 \times 0.24} = 1.33$

gain = 0.88 + 1.33 − 0

$= 2.22$ →

Splitting criteria = 8.87

cgpa < 8.87

−0.6, 0.4, −0.6, 0.4 | 0.4

Stage 2 model

cgpa < 7.62

output → −0.6, 0.4, −0.6 (−1.11) | 0.4, 0.4 (1.66) ←

$\frac{-0.6 + 0.4 - 0.6}{3 \times 0.6 \times 0.4} = \frac{-0.8}{3 \times 0.24} = -1.11$

$\frac{0.4 + 0.4}{2 \times 0.24} = 1.66$

output = $\frac{\sum \text{residual}_i}{\sum \text{prev\_prob}_i (1 - \text{prev\_prob}_i) + \lambda = 0}$

stage 1 → prediction

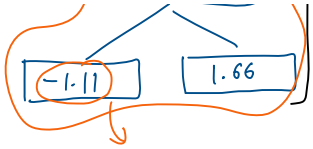
cgpa	placed?	pred1(lo)	pred1(prob)	res1	pred2(lo)	pred2(prob)	res 2
→ 5.70	0	0.405	0.6	−0.6	0.072	0.518	−0.51
→ 6.25	1	0.405	0.6	0.4	0.072	0.518	0.48
→ 7.10	0	0.405	0.6	−0.6	0.072	0.518	−0.51
→ 8.15	1	0.405	0.6	0.4	0.903	0.712	0.28
→ 9.60	1	0.405	0.6	0.4	0.903	0.712	0.28

Stage 2 model — Similar

$m1$ (log odds) + $0.3 \times m2$ (decision tree)

0.405 + 0.3 × [cgpa < 7.62 → −1.11 | 1.66]

8.15	1	0.405	0.6	0.4	0.903	0.712	0.28
9.60	1	0.405	0.6	0.4	0.903	0.712	0.28



$0.405 + 0.3 \times (-1.11)$

$0.405 + 0.3 \times (-1.11)$

$0.405 + 0.3 \times 1.66$

log(odds) → prob

$\frac{e^{\log(odds)}}{1 + e^{\log(odds)}}$

m3 → decision

same flow

Stage 3

$[m1 + 0.3 \times m2 + 0.3 \times m3]$ → pred → prob → res3

cgpa/res3 → my decision

next video

→ maths form → xgboost