

Prerequisite & Disclaimer

11 November 2023 09:29

{
→ Xgboost Reg
→ Gradient Boosting
→ log(odds)
} classifying

Disclaimers
→ optim
→ xgboost {
 perf
 speed
}
→ calculation → mistakes

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 \rightarrow classification

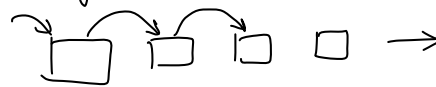
6.61 \rightarrow 1,0

\rightarrow overview
 \rightarrow step by step

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

xgboost \rightarrow GBPT \rightarrow

stage wise add



base
model
(mean
 $\log(\text{odds})$)

m_1

+

\log

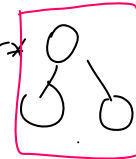


dt

m_2

+

\log



m_3

...

m_n

GBPT \rightarrow vanilla $\left\{ \begin{array}{l} \text{entropy} \\ \text{gini} \end{array} \right\}$
dt

xgboost \rightarrow diff $\left\{ \begin{array}{l} \text{similarity} \\ \text{score} \end{array} \right\}$
dt \rightarrow xgboost

Stage 1 - $\square \rightarrow \log(\text{odds}) \rightarrow \text{prob}$

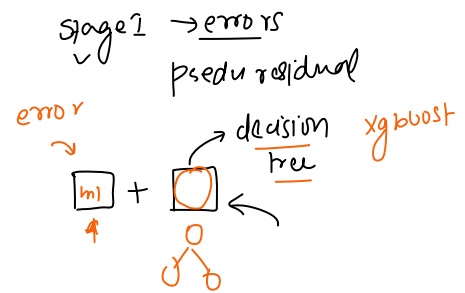
base estimator $\rightarrow \text{mean}$
 $\log(\text{odds}) \rightarrow \log\left(\frac{p}{1-p}\right)$ $p \rightarrow \text{prob +ve class (1)}$

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

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

cgpa	placed?	pred 1 (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

cgpa	placed?	pred 1 (lo)	pred 1 (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



split - splitting criteria \leftarrow

cgpa	res1
5.70	-0.6 \leftarrow
6.25	0.4
7.10	-0.6
8.15	0.4
9.60	0.4

leaf node \rightarrow $-0.6, 0.4, -0.6, 0.4, 0.4$ \rightarrow similarity score

$SS = 0$ \rightarrow maximize

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

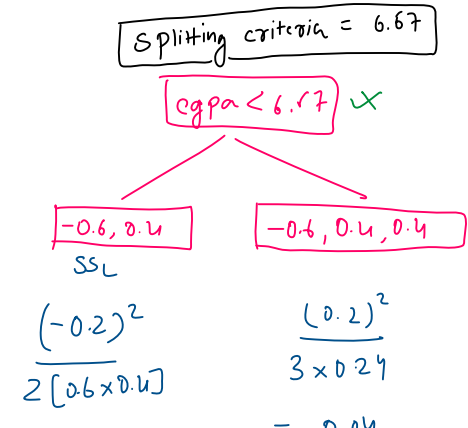
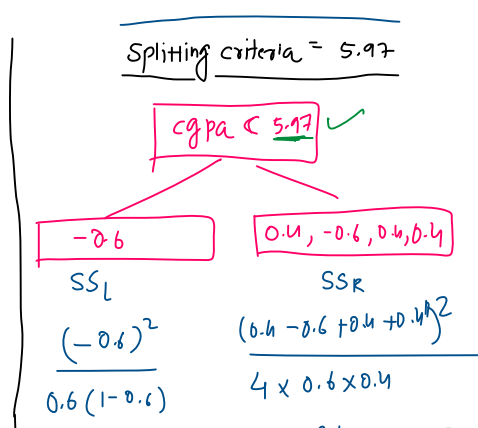
$\sum \text{prev_prob}_i (1 - \text{prev_prob}_i) + \lambda \rightarrow \text{reg } \lambda = 0$

$= (-0.6 + 0.4 - 0.6 + 0.4 + 0.4)^2 = 0$

$5 \times [0.6(1-0.6)] + 0$

gain \rightarrow 1.17

cgpa	res1
5.70	-0.6 \leftarrow
6.25	0.4
7.10	-0.6
8.15	0.4
9.60	0.4



$$\begin{array}{c|c} \text{cgpa} & \text{res1} \\ \hline 8.15 & 0.4 \\ 8.87 & 0.4 \\ 9.60 & 0.4 \end{array}$$

$$\begin{aligned} & \frac{(-0.6)^2}{0.6(1-0.6)} = \frac{0.36}{0.24} = 1.5 \\ & \frac{0.36}{4 \times 0.6 \times 0.4} = \frac{0.36}{0.96} = 0.37 \\ & \text{gain} = (SS_L + SS_R) - SS_{\text{root}} \\ & = (1.5 + 0.37) - 0 = 1.87 \end{aligned}$$

$$\begin{aligned} & \frac{0.04}{2 \times 0.24} = 0.08 \\ & \frac{0.04}{3 \times 0.24} = 0.05 \\ & \text{gain} = 0.08 + 0.05 - 0 = 0.13 \end{aligned}$$

splitting criteria = 7.62

splitting criteria = 8.87

cgpa	res1
5.70	-0.6
6.25	0.4
7.10	-0.6
8.15	0.4
9.60	0.4

$$\begin{aligned} & \text{cgpa} < 7.62 \\ & \begin{array}{l} \text{Left: } \frac{(-0.8)^2}{3 \times 0.24} = 0.88 \\ \text{Right: } \frac{(0.8)^2}{2 \times 0.24} = 1.33 \end{array} \\ & \text{gain} = 0.88 + 1.33 - 0 = 2.22 \end{aligned}$$

$$\begin{aligned} & \text{cgpa} < 8.87 \\ & \begin{array}{l} \text{Left: } \frac{(-0.6)^2}{3 \times 0.24} = 0.88 \\ \text{Right: } \frac{(0.4)^2}{2 \times 0.24} = 0.67 \end{array} \end{aligned}$$

stage 2 model

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

output

$$\begin{aligned} & \text{cgpa} < 7.62 \\ & \begin{array}{l} \text{Left: } \frac{(-0.6)^2}{3 \times 0.6 \times 0.4} = -0.8 \\ \text{Right: } \frac{(0.4)^2}{2 \times 0.24} = 0.67 \end{array} \\ & \text{stage 1} \rightarrow \text{prediction} \end{aligned}$$

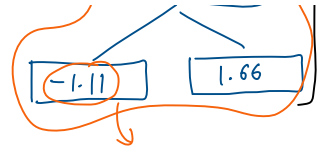
cgpa	placed?	pred1(lo)	pred1(prob)	res1	pred2(lo)	pred2(prob)	res2
5.70	0	0.405	0.6	-0.6	0.072	0.518	-0.51
6.25	1	0.405	0.6	0.4	0.072	0.518	0.48
7.10	0	0.405	0.6	-0.6	0.072	0.518	-0.51
8.15	1	0.405	0.6	0.4	0.903	0.712	0.28
9.60	1	0.405	0.6	0.4	0.903	0.712	0.28

Stage 2 model

Similar

$$\begin{aligned} & \overline{m1} + 0.3 \times \overline{m2} \\ & (0.405) + 0.3 \times (\text{decision tree}) \\ & \text{cgpa} < 7.62 \end{aligned}$$

→	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



$$0.405 + 0.3 \times (-1.11)$$

etc

$$0.405 + 0.3 \times (-1.11)$$

$$0.405 + 0.3 \times 1.66$$

log(odds)
→ prob

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

m3 → decision

same flow

Stage 3
 $[m1 + 0.3 \times m2 + 0.3 \times m3] \rightarrow \text{pred} \rightarrow \text{prob} \rightarrow \text{res3}$
 cgra/res3 → m4 → decision

next video

→ maths → xgbost
 form