

XGBoost for Classification

Problem:

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

⇒ Stage 1: Calculate log of odds of column 'placed'
 $\log \text{ of odds} = \log_e \left(\frac{P}{1-P} \right)$, $P \rightarrow$ Probability of +ve class (1)

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

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

	cgpa	placed	pred1(lo)	pred1(prob)	residual1 placed - pred1(prob)
① 5.97	5.70	0	0.405	0.60	-0.6
② 6.67	6.25	1	0.405	0.60	0.4
	7.10	0	0.405	0.60	-0.6
③ 7.62	8.15	1	0.405	0.60	0.4
④ 8.87	9.60	1	0.405	0.60	0.4

$$\text{Similarity Score (SS)} = \frac{(\sum \text{residuals})^2}{\# \text{residuals} \times \text{prev-prob}(1 - \text{prev-prob}) + \lambda}$$

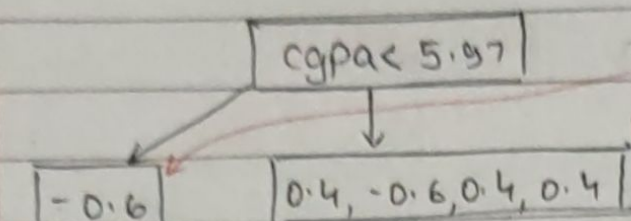
reg. parameter
assume = 0

Node

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

⇒ Sort cgpa column then take avg of adjacent cgpa values for splitting criteria.

① Splitting criteria = 5.97



Residuals of corresponding cgpa for the given condition

$$SS_L = \frac{(-0.6)^2}{0.6 \times (1-0.6)}$$

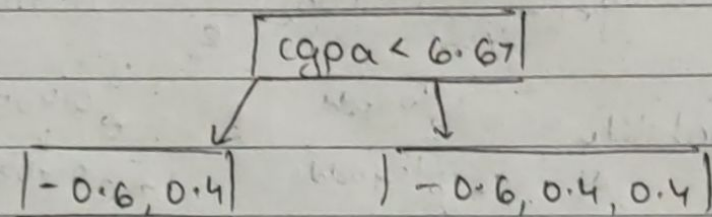
$$SS_R = \frac{(0.4 - 0.6 + 0.4 + 0.4)^2}{4 \times 0.6 \times 0.4}$$

$$SS_L = 1.5$$

$$SS_R = 0.37$$

$$\text{gain} = SS_L + SS_R - SS_{\text{Parent}} \\ = 1.5 + 0.37 - 0 = 1.87$$

② Splitting criteria = 6.67

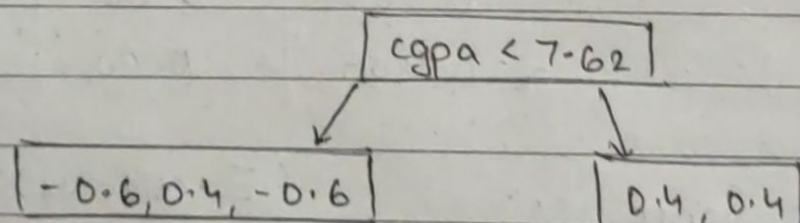


$$SS_L = \frac{(-0.6 + 0.4)^2}{2 \times 0.6 \times 0.4} = 0.08$$

$$SS_R = \frac{(-0.6 + 0.4 + 0.4)^2}{3 \times 0.6 \times 0.4} = 0.05$$

$$\text{gain} = 0.08 + 0.05 - 0 = 0.13$$

③ Splitting criteria = 7.62



$$SS_L = \frac{(-0.6 + 0.4 - 0.6)^2}{3 \times 0.6 \times 0.4} = 0.88$$

$$SS_R = \frac{(0.4 + 0.4)^2}{2 \times 0.6 \times 0.4} = 1.33$$

$$\text{gain} = 0.88 + 1.33 - 0 = 2.22 \rightarrow \text{Highest}$$

Output of leaf nodes =

\sum residuals

#residuals \times prev-prob \times (1-prev-prob)

⇒

$$cgpa < 7.62$$

$$[-0.6, 0.4, -0.6]$$

$$[0.4, 0.4]$$

$$o/p = \frac{-0.6 + 0.4 - 0.6}{3 \times 0.6 \times 0.4}$$

$$= -1.11$$

$$o/p = \frac{0.4 + 0.4}{2 \times 0.6 \times 0.4}$$

$$= 1.66$$

⇒

$$\text{Stage 2: } 0.405 + 0.3 \times$$

learning rate
(η) eta

$$cgpa < 7.62$$

$$[-1.11]$$

$$[1.66]$$

placed - pred2
(prob)

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

$$0.405 + 0.3 \times (1.66)$$

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

$$\text{prob} = \frac{e^{\log odds}}{1 + e^{\log odds}}$$

Repeat by adding more stages & decision trees until residual $\rightarrow 0$