

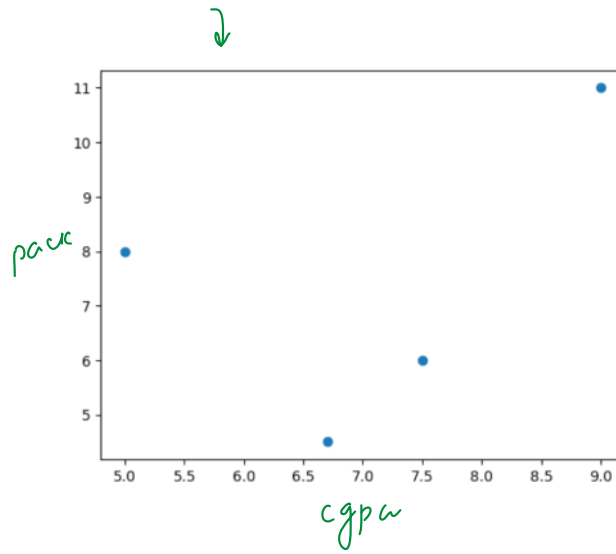
The Problem Statement

23 October 2023 15:05

lpa

cgpa	package
6.7	4.5
9.0	11.0
7.5	6.0
5.0	8.0

⑦



→ regression

new → cgpa → xgboost
↓
package

Solution Overview

23 October 2023 15:05

pseudo-residual

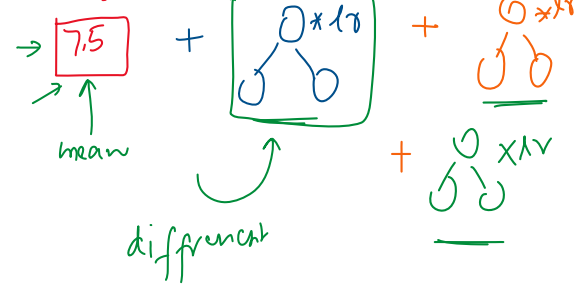
[perf / speed]

4.1, 9.1

xgboost
↳ gradient
boosting

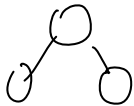
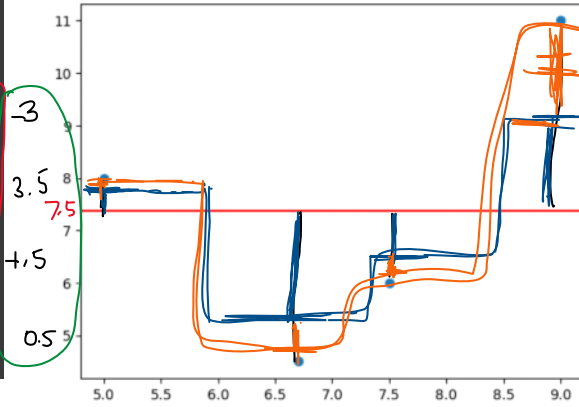
base estimator → mean

stage 1 [stage II]



feature

cgpa	package
6.7	4.5
9.0	11.0
7.5	6.0
5.0	8.0



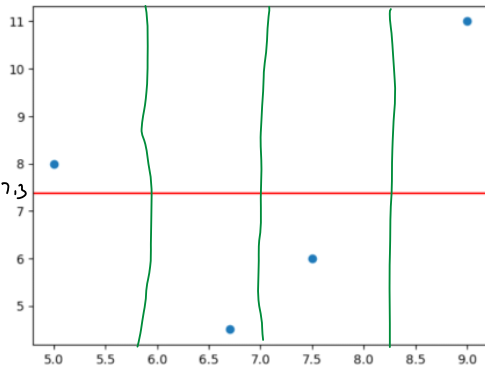
Step by Step Calculation

23 October 2023 15:07

7.3

stage 1 → mean

cgpa	package	model 1	residual 1
6.7	4.5	7.3	-2.8
9.0	11.0	7.3	3.7
7.5	6.0	7.3	-1.3
5.0	8.0	7.3	0.7



cgpa | residual 1

decision tree

→ 0.02

→ Similarity score

$SS = \frac{(\text{sum of residuals})^2}{\# \text{residuals} + 1}$ → reg parameter

$\lambda = 0$

$\frac{(-2.8 + 3.7 - 1.3 + 0.7)^2}{4 + 0} = \frac{(0.3)^2}{4} = 0.02$

5.0 } → 5.85
6.7 } → 7.1
7.5 } → 8.25
9.0 }

cgpa	residual 1
5.0	0.7
6.7	-2.8
7.5	-1.3
9.0	3.7

Splitting criteria ① → 5.85

→ $cgpa < 5.85$ ✓

$SS_L = \frac{(0.7)^2}{1 + 0} = 0.49$

$SS_R = \frac{(-2.8 - 1.3 + 3.7)^2}{3} = \frac{0.16}{3} = 0.05$

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

= $0.49 + 0.05 - 0.02$

= 0.52 ✓

Splitting criteria ② → 7.1

→ $cgpa < 7.1$ → Similar

$SS_L = \frac{(0.7 - 2.8)^2}{2} = \frac{4.41}{2} = 2.20$

$SS_R = \frac{(-1.3 + 3.7)^2}{2} = \frac{5.76}{2} = 2.88$

gain = $SS_L + SS_R - SS_{\text{root}}$

= $2.20 + 2.88 - 0.02$

= 5.06 ✓

Splitting Criteria ③ → 8.25

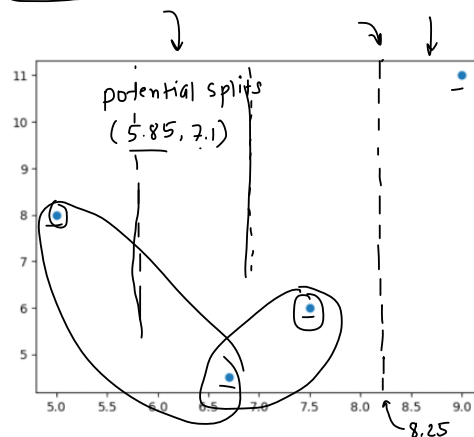
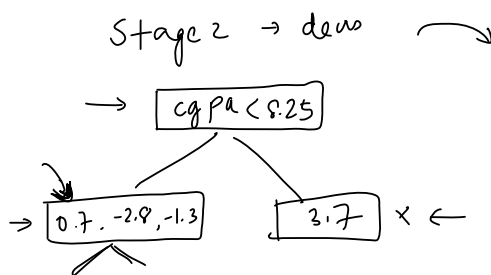
→ $cgpa < 8.25$ ✓

$SS_L = \frac{(0.7 - 2.8 - 1.3)^2}{3} = \frac{11.56}{3} = 3.85$

$SS_R = \frac{(3.7)^2}{1} = 13.69$

gain = $3.85 + 13.69 - 0.02$

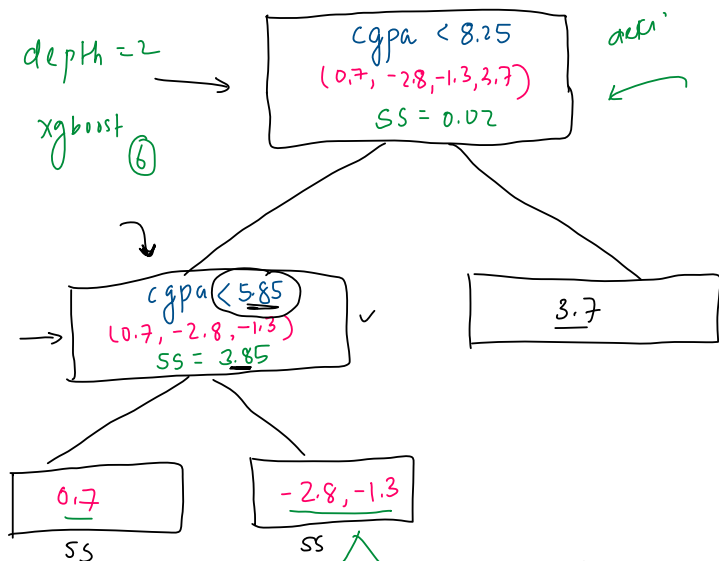
= 17.52



① Splitting Criteria $\rightarrow 5.85$

depth = 2

xgboost ⑥

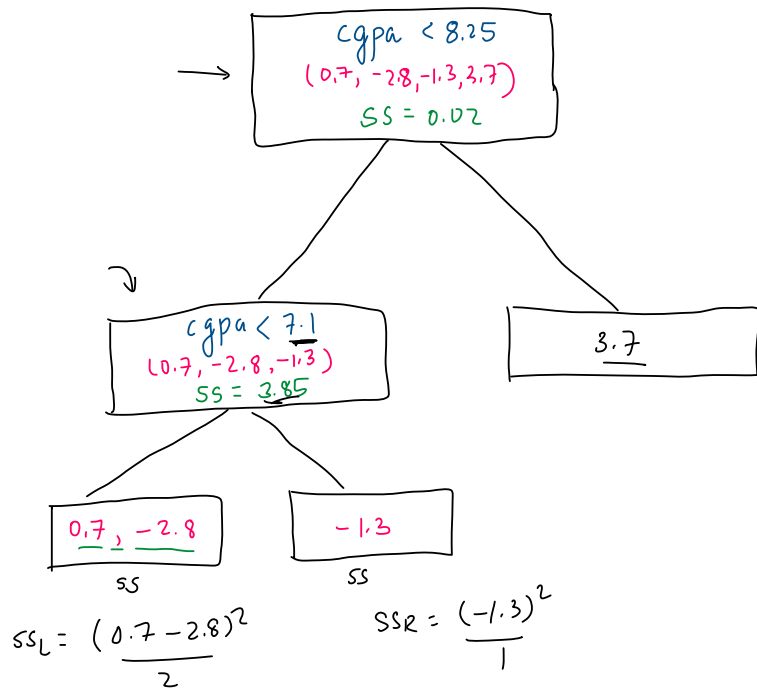


$$SS_L = \frac{(0.7)^2}{1} = 0.49$$

$$SS_R = \frac{(-2.8 - 1.3)^2}{2} = \frac{(-4.1)^2}{2} = 8.40$$

$$\text{gain} = (SS_L + SS_R - SS_P) = 0.49 + 8.40 - 3.85 = 5.04 \checkmark$$

② Splitting Criteria $\rightarrow 7.1$

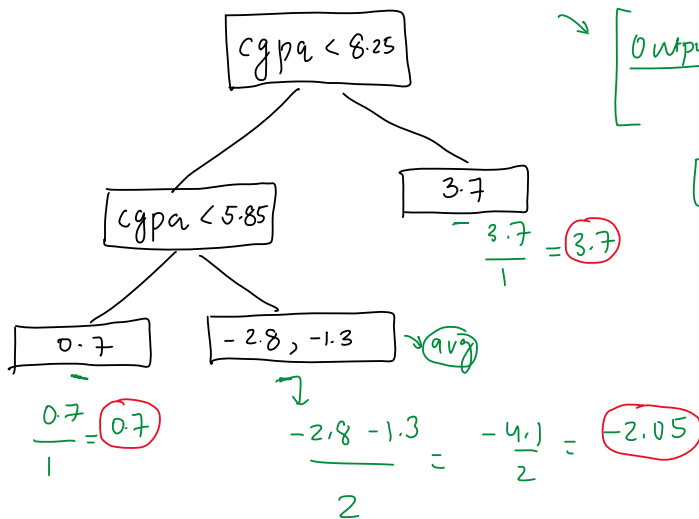


$$SS_L = \frac{(0.7 - 2.8)^2}{2} = \frac{(-2.1)^2}{2} = 2.20$$

$$SS_R = \frac{(-1.3)^2}{1} = 1.69$$

$$\text{gain} = 2.20 + 1.69 - 3.85 = 0.04 - \times$$

The final decision tree



$$\text{Output} = \frac{\text{sum of residuals}}{\# \text{residuals} + \lambda}$$

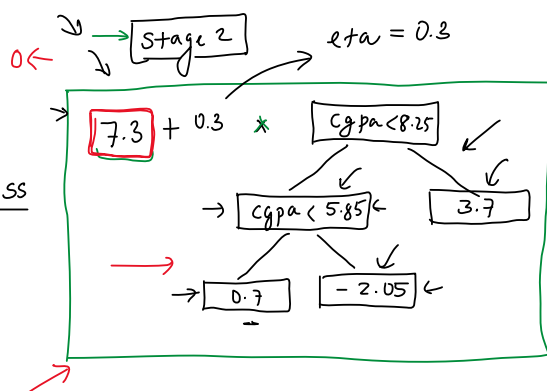
$$\lambda = 0 \rightarrow \lambda = 0 \rightarrow \text{avg}$$

	cgpa	package	model 1	residual 1	model 2	residual 2
→	6.7	4.5	7.3	-2.8	6.69	-2.19
→	9.0	11.0	7.3	3.7	8.41	2.59
→	7.5	6.0	7.3	-1.3	6.69	-0.69
→	5.0	8.0	7.3	0.7	7.51	0.49

Stage 3

1.12

$$7.3 + 0.3 \times (-2.05)$$



Stage (3)
 \swarrow
 cgpa | res 2

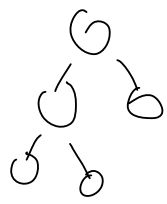
mean + $\lambda \times dt1 + \lambda \times dt2$

$7.3 + 0.3 \times (-2.05)$
 $7.3 + 0.3 \times 3.7$
 $7.3 + 0.3 \times (-2.05)$
 $7.3 + 0.3 \times 0.7$

100/200 \rightarrow residual $\rightarrow 0$

category \rightarrow binary
 \rightarrow multipliers

cgpa | package \rightarrow $\begin{bmatrix} f_1 \\ \text{cgpa} \end{bmatrix}$ | $\begin{bmatrix} f_2 \\ \text{12th mark} \end{bmatrix}$ | package



$\left\{ \begin{array}{l} ss = \\ out \end{array} \right.$

$\lambda \rightarrow$

\rightarrow

xgboost