Q 11.6

$\lambda = 0.5$, $r = 0.5$, maximum depth = 2

data;

y	8	3	6	7
x	1	2	6	6

a) w/ $f_0 = 0.5$;

y	x	f_0	r
8	1	0.5	7.5
3	2	0.5	2.5
6	6	0.5	5.5
7	6	0.5	6.5

$$\text{similarity} = 0.5 \times \frac{(\text{sum of residuals})^2}{\#\text{ of residuals} + r}$$

consider 3 possible splits; $x < 1.5$, $x < 2.5$, $x < 6.5$

$$\text{similarity of the root} = \frac{1}{2} \times \frac{(7.5 + 2.5 + 5.5 + 6.5)^2}{4 + 0.5} = \frac{1}{2} \times \frac{22^2}{4.5} = \frac{1}{2} \times \frac{968}{9} = \frac{484}{9}$$

- $x < 1.5$; left - $y = \{8\}$

$$\Rightarrow sim_{left} = \frac{1}{2} \times \frac{7.5^2}{1.5} = \frac{75}{4}$$

right - $y = \{3, 6, 7\}$

$$\Rightarrow sim_{right} = \frac{1}{2} \times \frac{(2.5 + 5.5 + 6.5)^2}{3.5} = \frac{841}{28}$$

$$\therefore \text{gain} = \left(\frac{75}{4} + \frac{841}{28}\right) - \frac{484}{9} = -\frac{629}{126}$$

- $x < 2.5$; left - $y = \{8, 3\}$

$$\Rightarrow sim_{left} = \frac{1}{2} \times \frac{(7.5 + 2.5)^2}{2.5} = 20$$

right - $y = \{6, 7\}$

$$\Rightarrow sim_{right} = \frac{1}{2} \times \frac{(5.5 + 6.5)^2}{2.5} = \frac{144}{5}$$

$$\therefore \text{gain} = \left(20 + \frac{144}{5}\right) - \frac{484}{9} = -\frac{224}{45}$$

- $x < 6.5$; left - $y = \{8, 3, 6, 7\}$

$$\Rightarrow sim_{left} = \frac{484}{9}$$

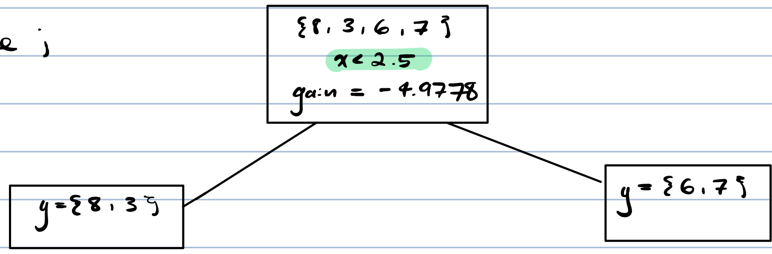
right - $y = \emptyset$ $\Rightarrow sim_{right} = 0$

$\therefore$ gain = 0, since we've basically just moving obs. down.

since $\text{gain}_{x<2.5} = -\frac{224}{45} > \text{gain}_{x<1.5} = -\frac{629}{126}$, we choose $x < 2.5$ as the root node split.

(≈ -4.9778) (≈ -4.992)

so, we currently have ;



we have no further splits to consider on the right, as both values correspond to $x = 6$.
so, we consider the last remaining split ; $x < 1.5$

· $x < 1.5$; left ; $y = \{8\}$

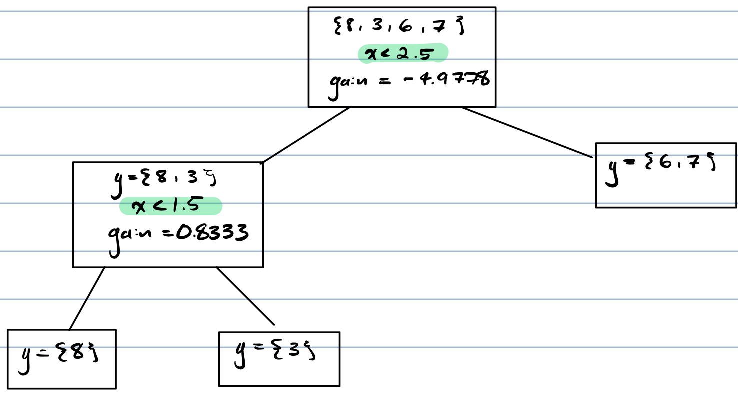
$$sim_{left} = \frac{1}{2} \times \frac{7.5^2}{1.5} = \frac{75}{4}$$

right ; $y = \{3\}$

$$sim_{right} = \frac{1}{2} \times \frac{2.5^2}{1.5} = \frac{25}{12}$$

$$\therefore \text{gain} = \left(\frac{75}{4} + \frac{25}{12}\right) - 20 = \frac{5}{6} \approx 0.8333$$

hence ,

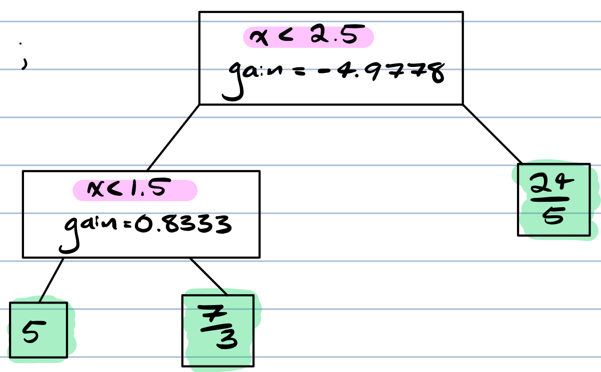


now, need to calculate fitted values.
from L → R ; $Y_{T_1} = \frac{7.5}{1.5} = 5$

$$Y_{T_2} = \frac{3.5}{1.5} = \frac{7}{3}$$

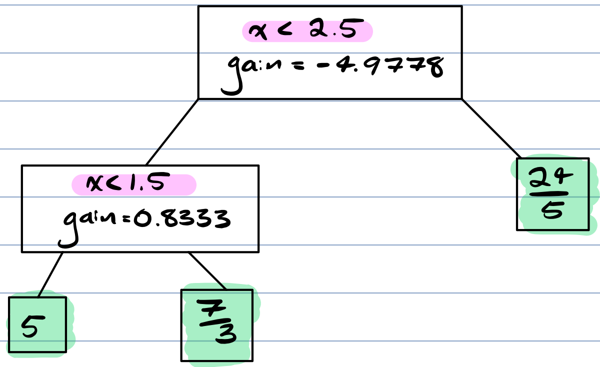
$$Y_{T_3} = \frac{5.5 + 6.5}{2.5} = \frac{24}{5}$$

$\therefore$ the first tree is ;



b) since we have $\alpha = 0.5$,
we would not trim the $x < 1.5$ split, since $0.8333 > 0.5$
and, as a result, we wouldn't trim the parent node.
so, we'll have the same tree.

$\therefore$ pruned tree:



c) ... already did this...

d) $f_1 \mid x < 1.5 = \overset{0.5}{f_0} + 0.5 \times 5 = 3$

$f_1 \mid 1.5 < x < 2.5 = 0.5 + 0.5 \times \frac{7}{3} = \frac{5}{3}$

$f_1 \mid x > 2.5 = 0.5 + 0.5 \times \frac{24}{5} = \frac{29}{10}$

e) 0.5 is nowhere close to any of the y-values, leading to an inefficient use of the XGboost algorithm. the residuals aren't saying much. hence, $\bar{y}$ would be a better choice, as the residuals would actually mean something when creating the tree & computing the gain.