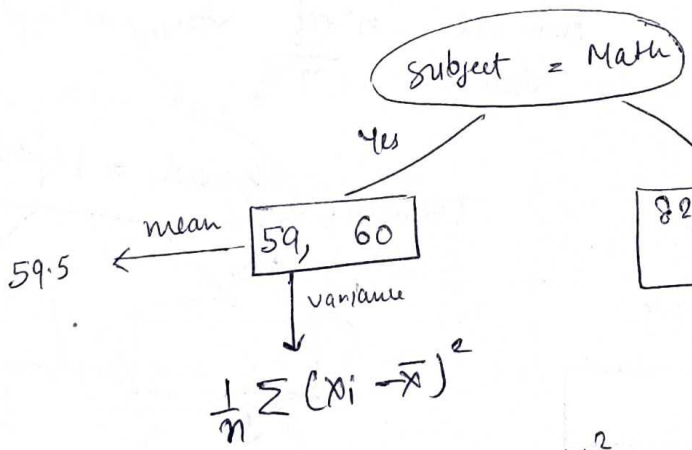
Cart for Regression

	Subject	Grade_level	Hour_studied	Test_Score
0	Math	Freshman	4	59
1	Physics	fresh man	1	82
2	Physics	Freshman	4	81
3	Math	Junior	6	60
4	Physics	Sophomore	1	73
5	Physics	Junior	3	85
6	Physics	Junior	4	61
7	Physics	Freshman	9	78
	binary ↑	multi class ↑	Numerical ↑	



$$\frac{1}{n}\sum (x_i - \bar{x})^2$$

$$\text{Variance} = \frac{(59 - 59.5)^2 + (60 - 59.5)^2}{2}$$

$$\frac{(82-75)^2 + (81-75)^2 + \cdots + (78-75)^2}{6}$$

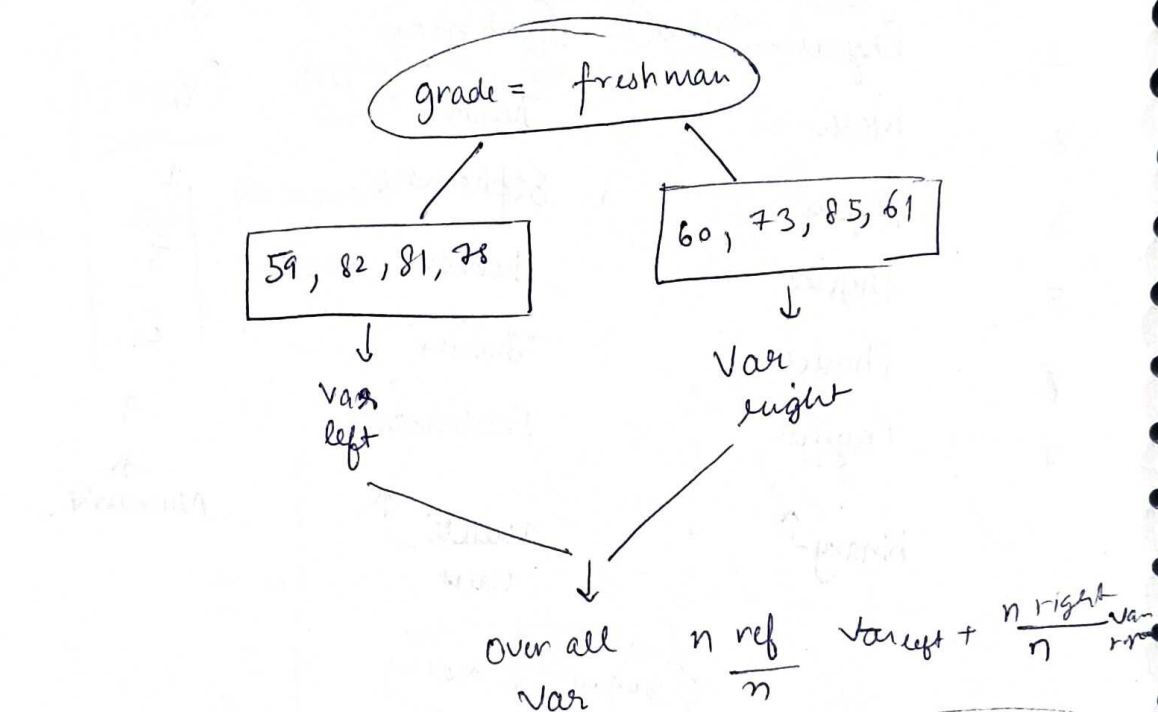
$$\text{Overall var} = \frac{2}{8} \times 2 + \frac{6}{8} \times 10 = 15$$

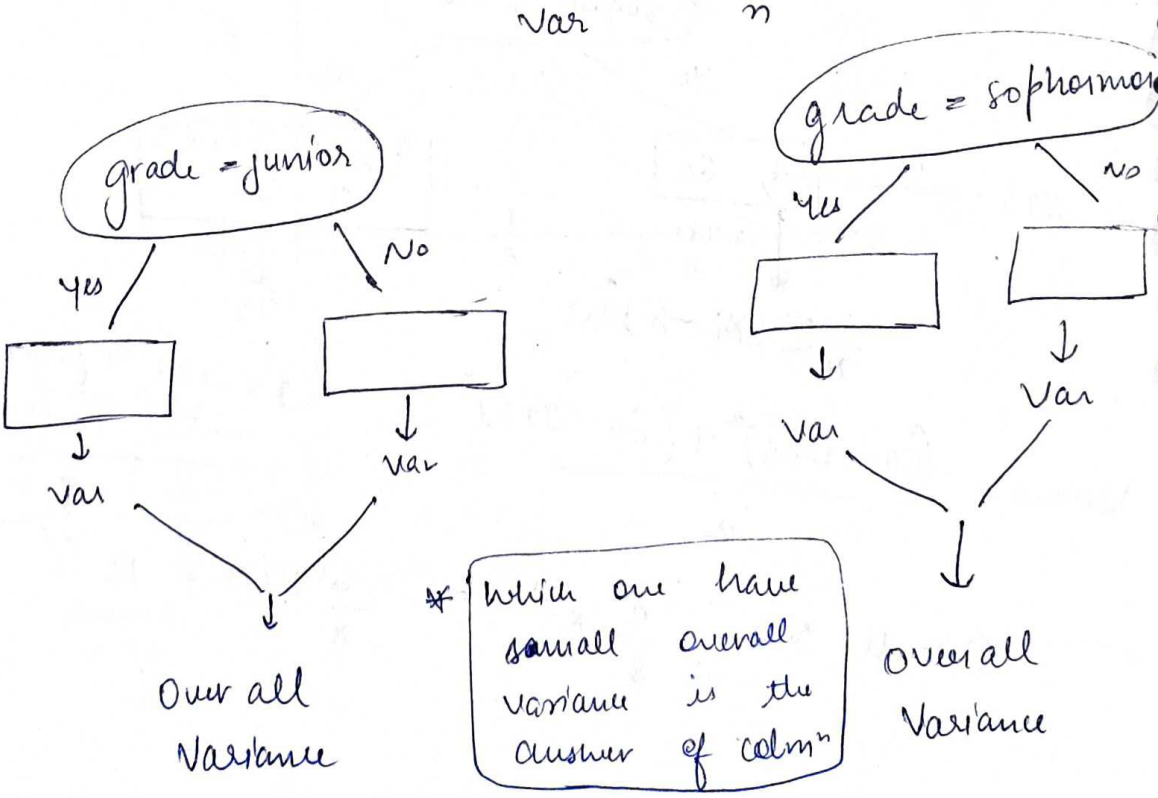
For multi class → 3 splits

{ freshman (junior , Sophomore) }

{ junior , (freshman , sophomore) }

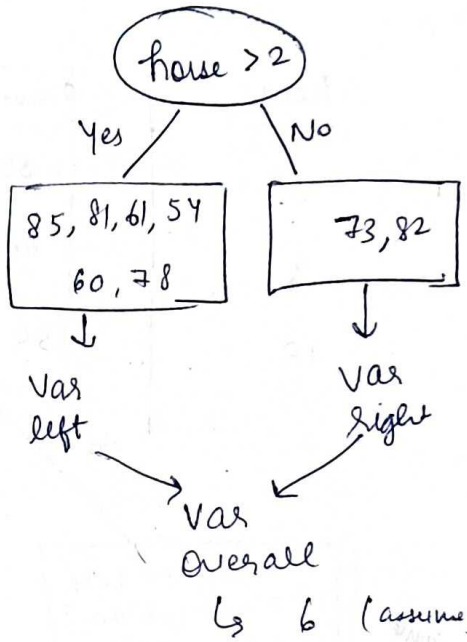
{ sophomore , (freshman , junior) }

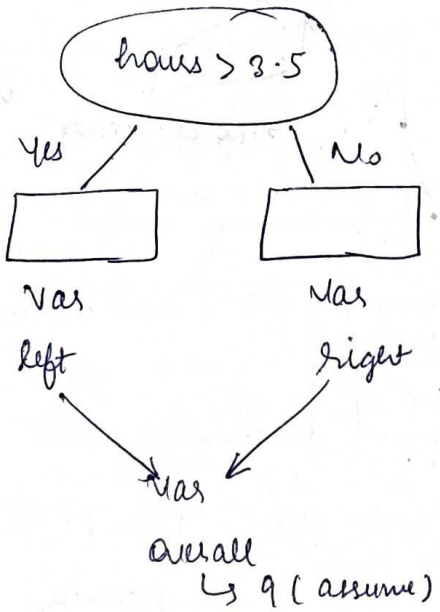


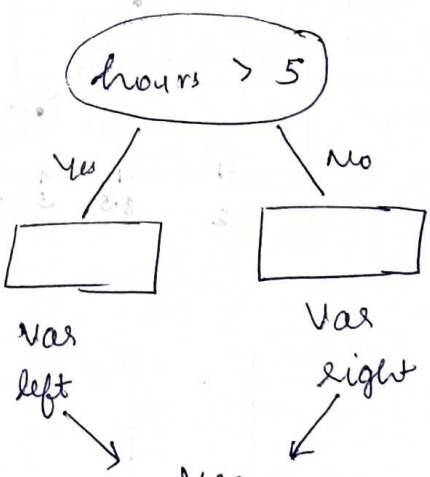


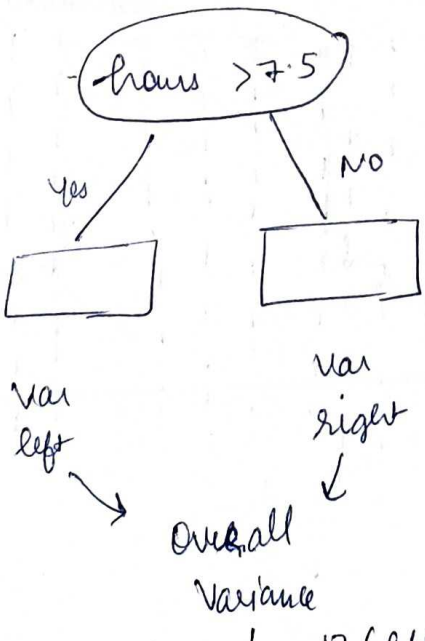
* which one have samall overall variance is the answer of colm

	1	82
	1	73
2	- - -	- - -
	3	85
3.5	- - -	- - -
	4	81
	4	61
	4	54
5	- - -	- - -
	6	60
7.5	- - -	- - -
	9	78







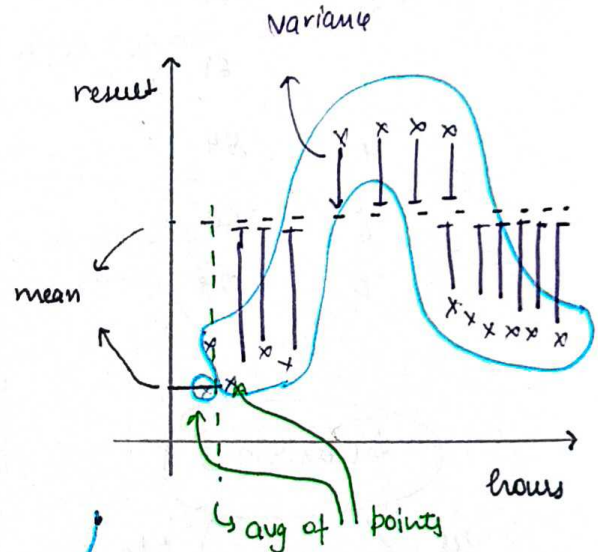


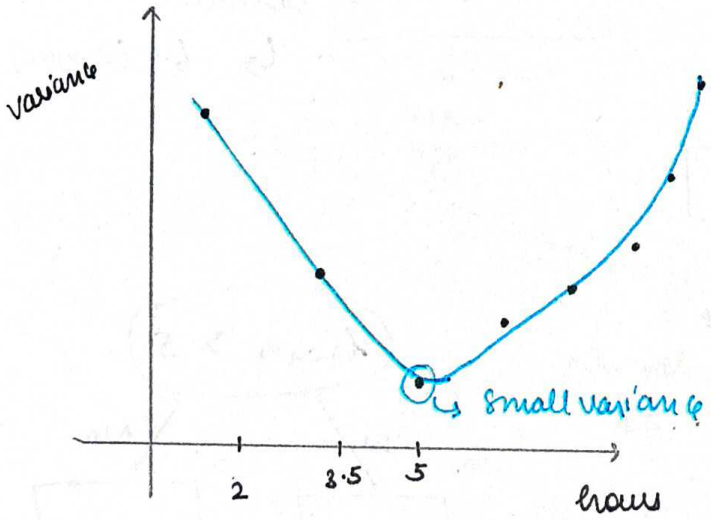
* Or we can use use MAE instead of overall and Var right and Var left.

Geometric Intution

hours	exam-result
5	50
6	60
10	30

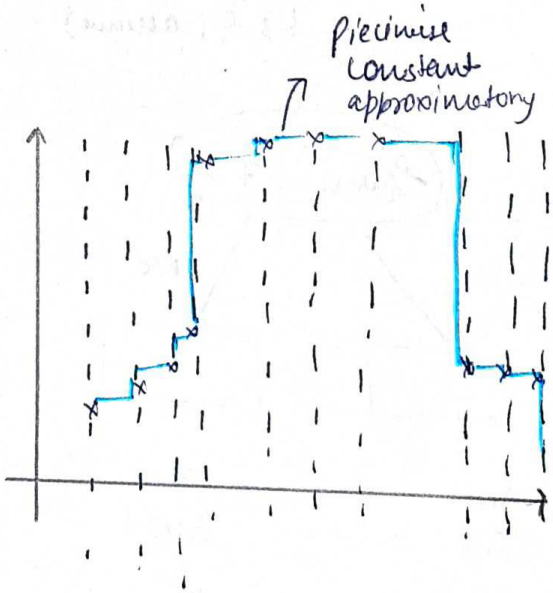
$V_{left} = 0$ $V_{right} = -$



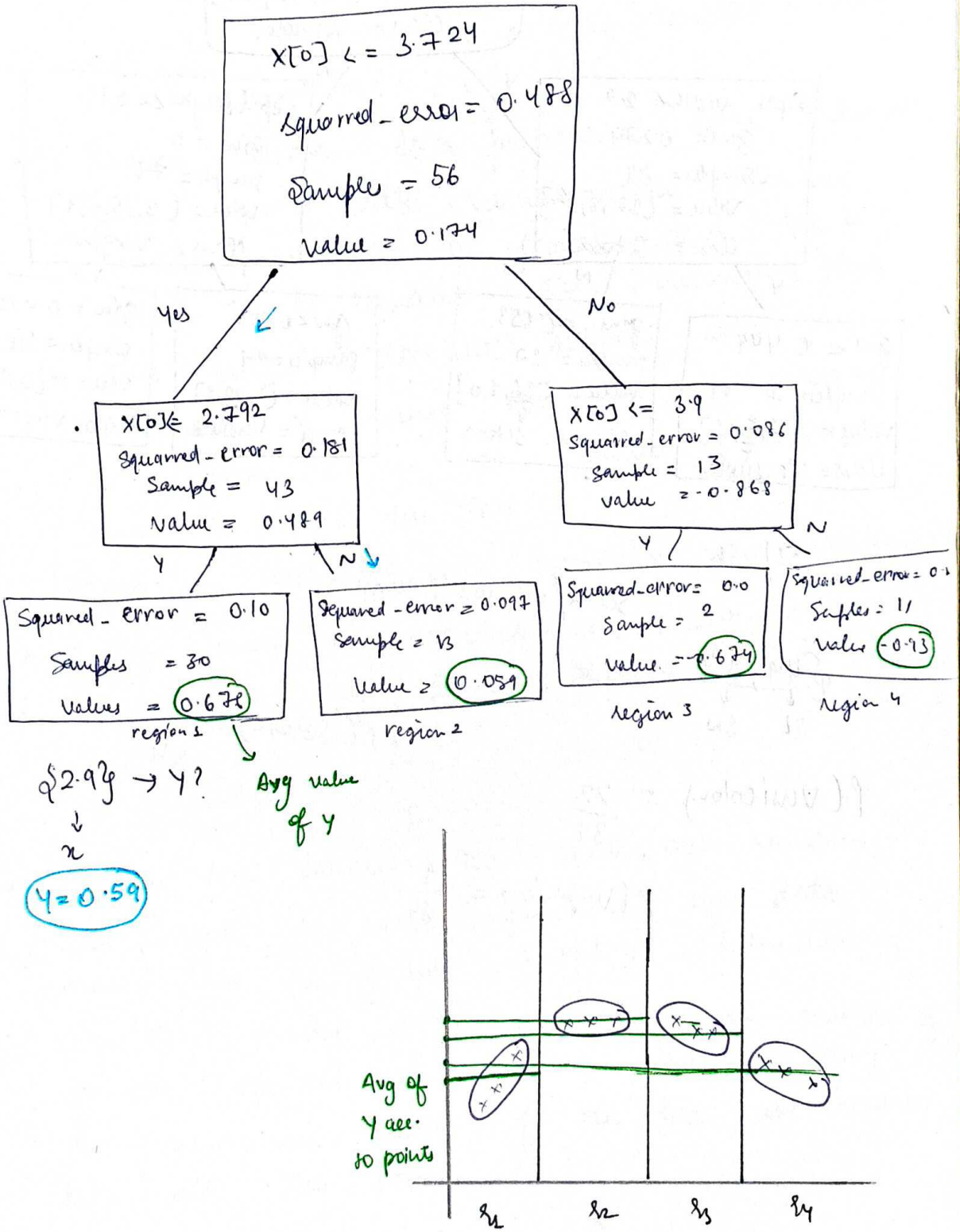


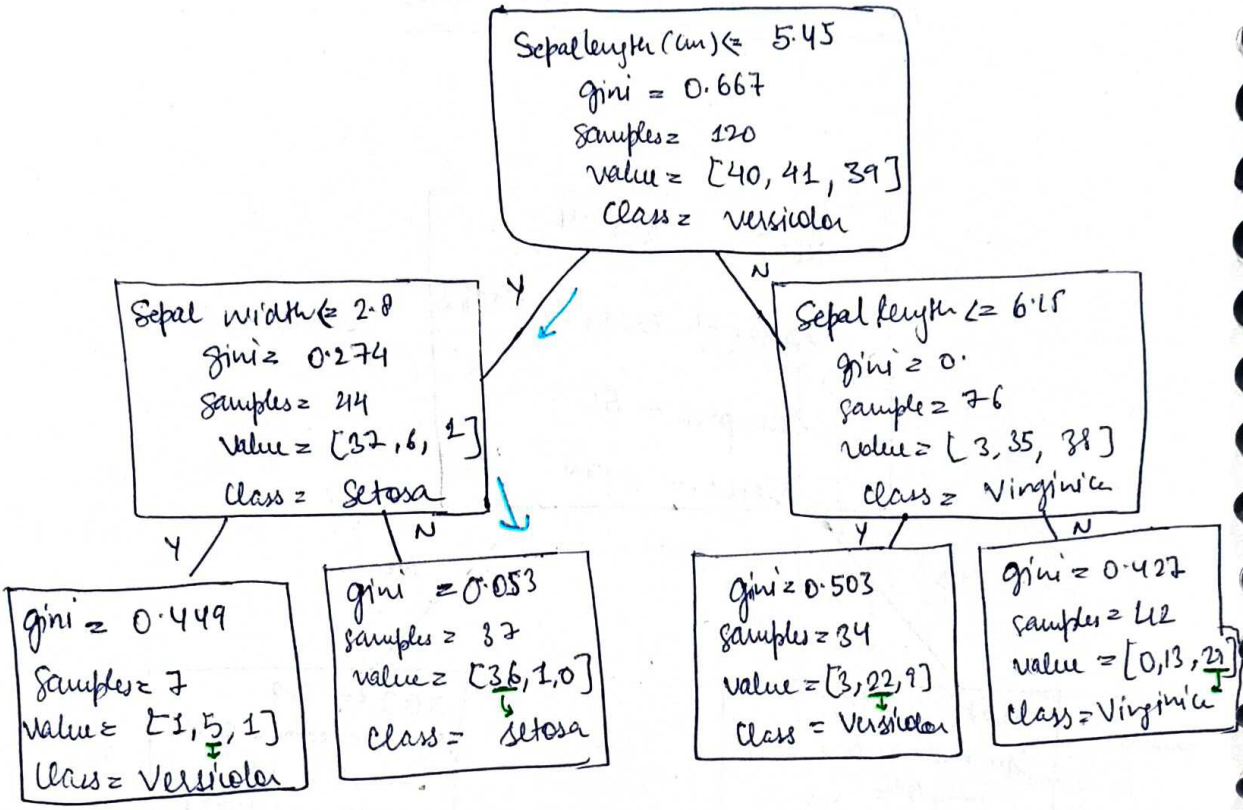
Touch every point

* High Overfitting



How Prediction is Done





SL / SW

Setosa / versicolor / verginica

{4.9, 3} ⇒ Setosa

SL SW

$P(\text{Setosa}) = \frac{3}{34}$

$P(\text{versicolor}) = \frac{22}{34}$

$P(\text{Virginica}) = \frac{1}{34}$

Advantages & Disadvantages

Advantages

- Simple to understand and to interpret. Trees can be visualized.
- Requires little data preparation. Other techniques often required data normalization, dummy variables need to be created and blank values to be removed. Note however that this module does not support missing values.
- The cost of using the tree (i.e predict data) is logarithmic in the number of data points used to train the tree.
- Able to handle both numerical and categorical
- can work on non-linear datasets
- can give you feature importance.

Disadvantages

- Decision tree learners can create over-complex trees that do not generalize the data well. This is called overfitting. Mechanism such as pruning setting the minimum number of samples required at a leaf node or setting the maximum depth of the tree are necessary to avoid this problem.
- Decision Tree can be unstable because small variation in the data might result in a completely different tree being generated. This

problem is mitigated by using decision trees within an ensemble.

- Prediction of decision tree are neither smooth nor continous, but piecewise constant approximations as seen in above figure. Therefore, they are not good at extrapolation. This limitation is inherent to the structure of decision tree models. They are very useful for interpretability and for handling non-linear relationship within the range of the training data, but they aren't designed for extrapolation. If extrapolation is imp. for your task, you might need to consider other types of model.