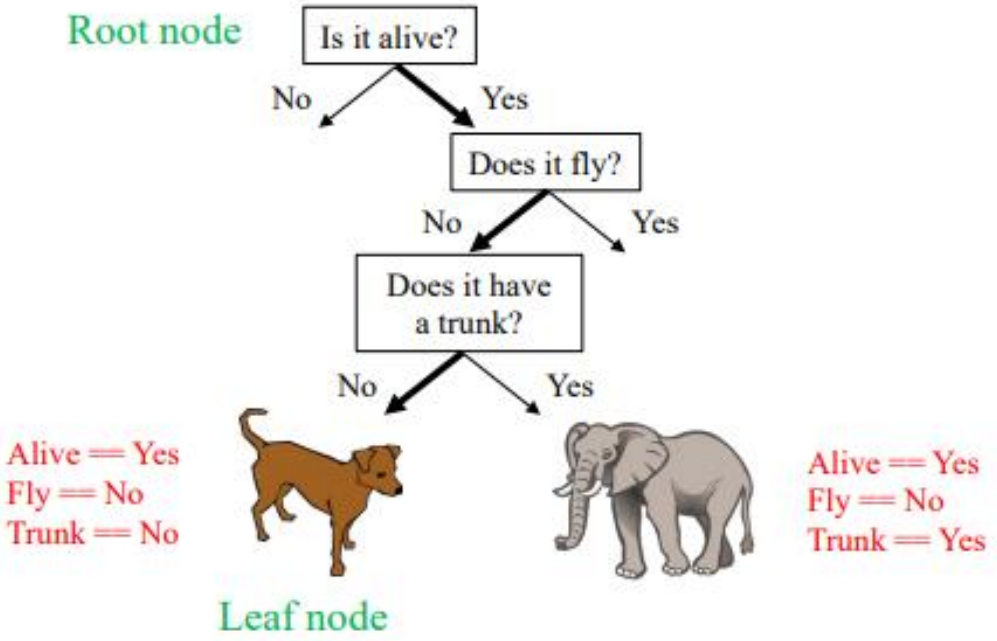
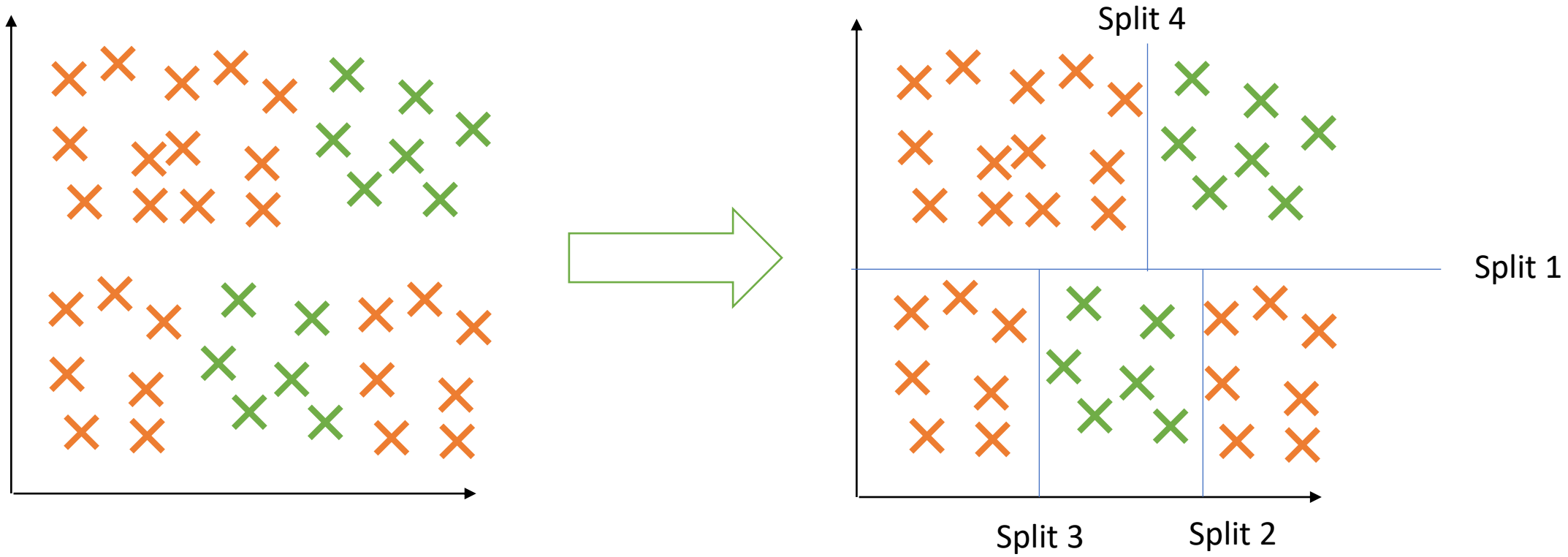
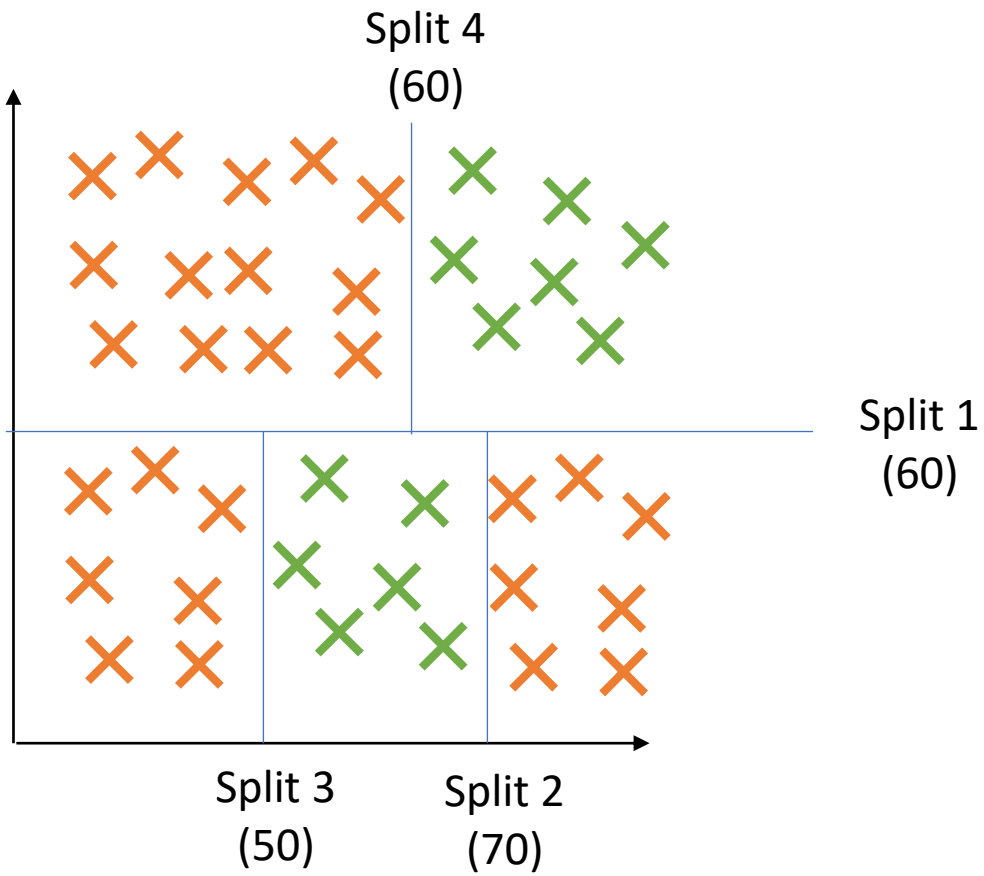
Decision Trees(CART – Classification and Regression Trees): -

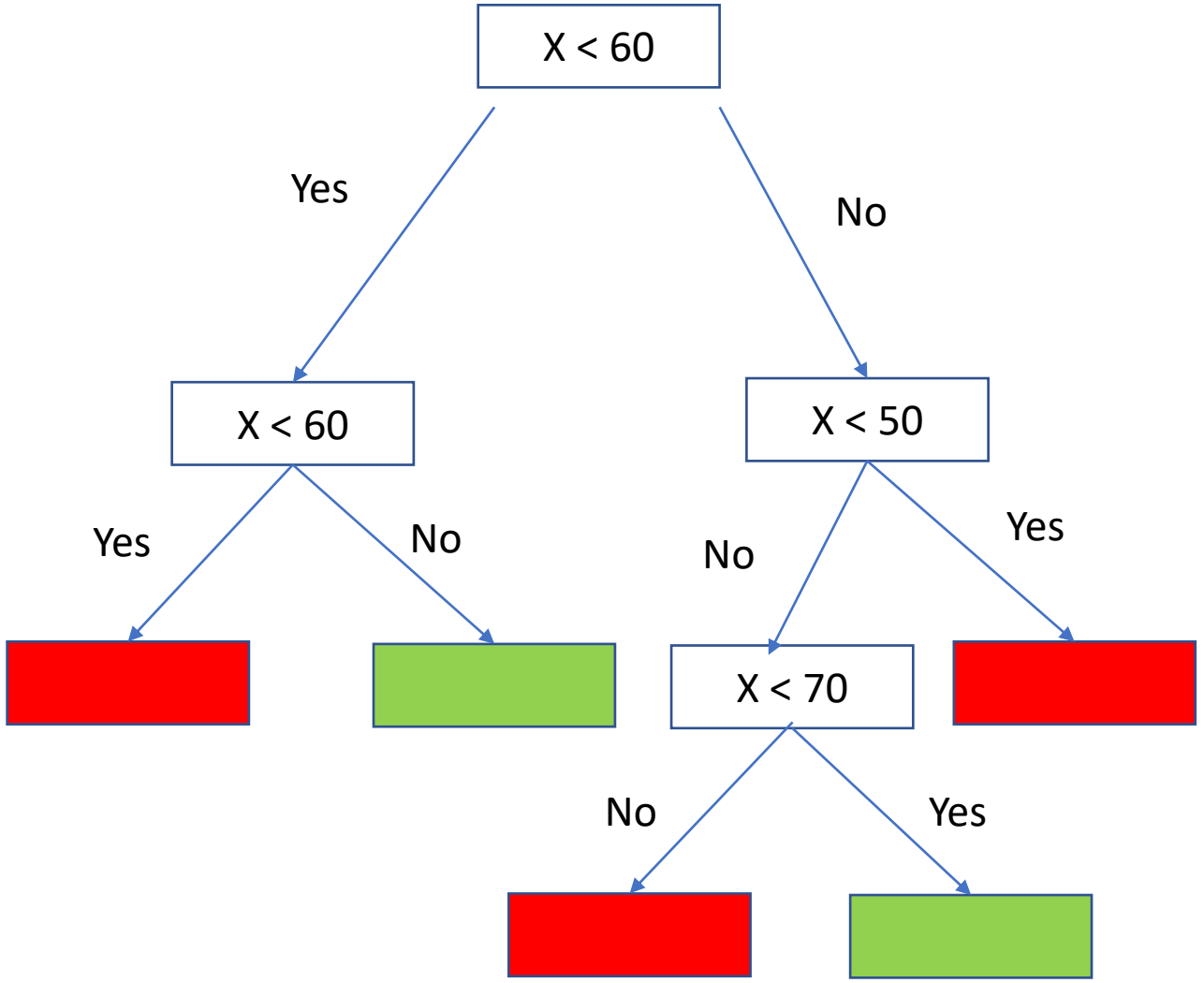


THARVAGURU



THARVAGURU

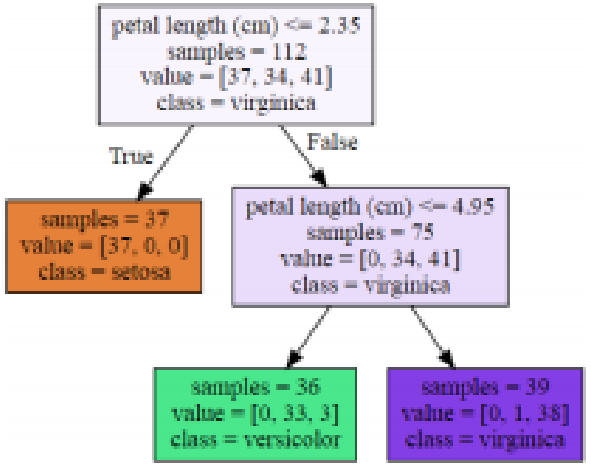

Split 4
(60)
Split 1
(60)
Split 3
(50)
Split 2
(70)


X < 60
Yes
No
X < 60
X < 50
Yes
No
No
Yes
X < 70
No
Yes

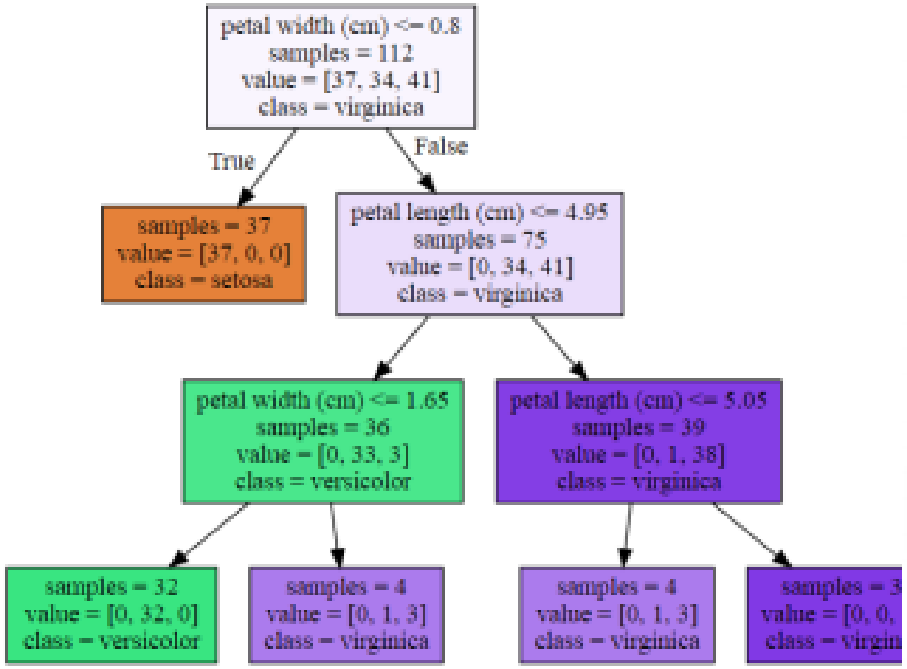
THARVAGURU

- max_depth: controls maximum depth (number of split points). Most common way to reduce tree complexity and overfitting.
- min_samples_leaf: threshold for the minimum # of data instances a leaf can have to avoid further splitting.
- In practice, adjusting only one of these (e.g. max_depth) is enough to reduce overfitting.
- Entropy and Gini are two different criterions used in a decision tree.

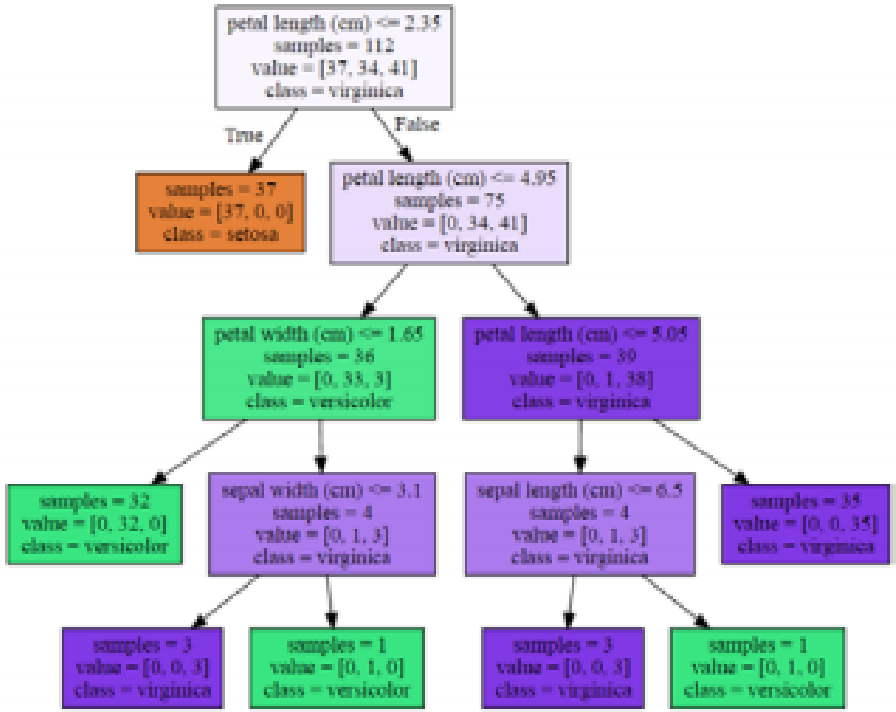
max_depth = 2



max_depth = 3



max_depth = 4



THARVAGURU