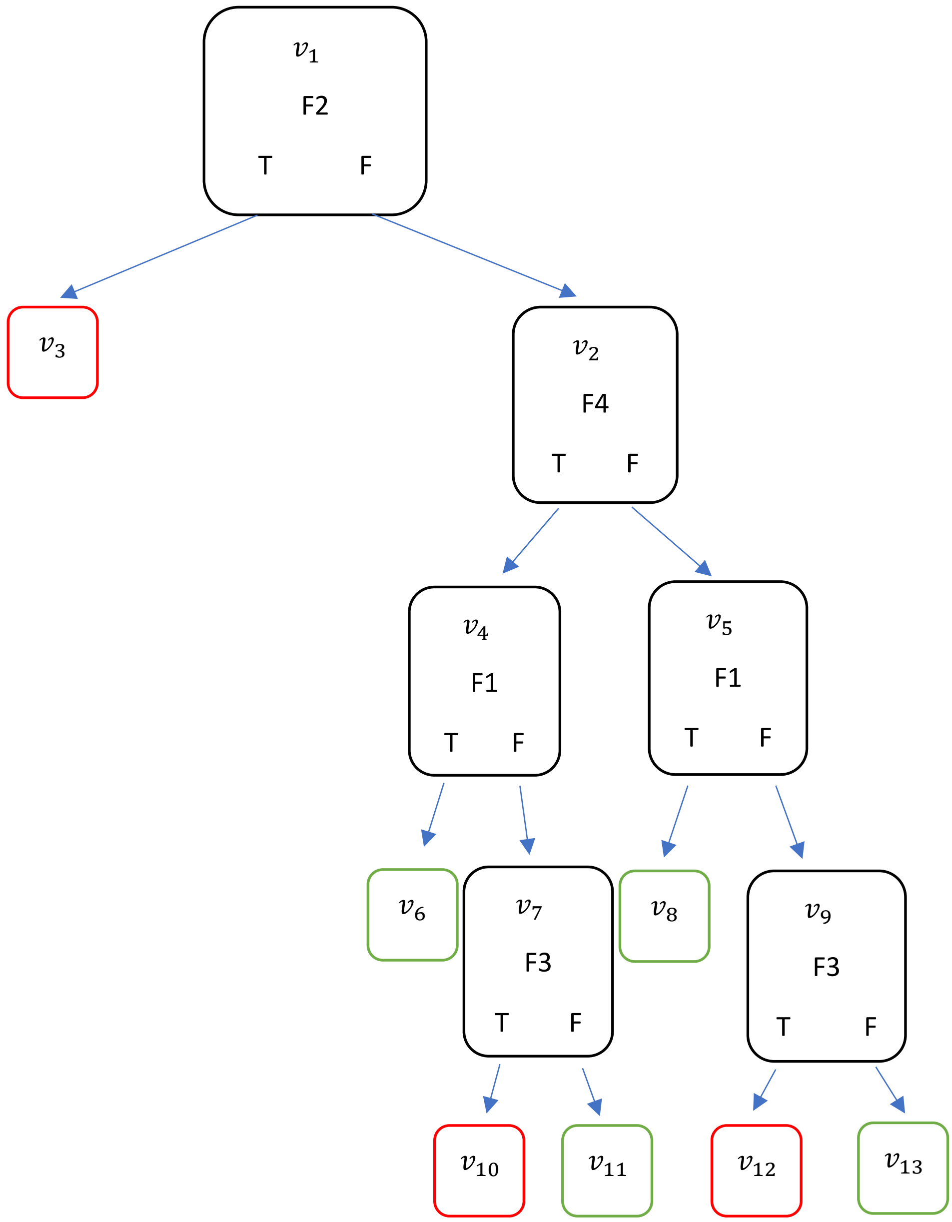
Short Homework 2 – Gal Kaptsenel 209404409

Q1

#	F1	F2	F3	F4	LABEL
1	T	T	F	T	F
2	T	T	F	F	F
3	F	F	T	T	F
4	F	F	T	F	F
5	T	F	T	T	T
6	T	F	T	F	T
7	F	F	F	T	T
8	F	F	F	F	T

ID3



v_1-

$$H(v_1) = -\frac{4}{8}\log_2\frac{4}{8} - \frac{4}{8}\log_2\frac{4}{8} = -\log_2\frac{4}{8} = -(\log_2 4 - \log_2 8) = -(2-3) = 1$$

$$H\left(v_{1_{F1=T}}\right) = H\left(v_{1_{F1=F}}\right) = -\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4} = -\log_2\frac{2}{4} = -(1-2) = 1$$

$$H\left(v_{1_{F2=T}}\right) = -\frac{0}{2}\log_2 0 - \frac{2}{2}\log_2\frac{2}{2} = 0$$

$$H\left(v_{1_{F2=F}}\right) = -\frac{4}{6}\log_2\frac{4}{6} = -\frac{2}{3}(1-\log_2 3)$$

$$H\left(v_{1_{F3=T}}\right) = H\left(v_{1_{F3=F}}\right) = -\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4} = -\log_2\frac{2}{4} = -(1-2) = 1$$

$$H\left(v_{1_{F4=T}}\right) = H\left(v_{1_{F4=F}}\right) = -\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4} = -\log_2\frac{2}{4} = -(1-2) = 1$$

$$IG(v_1, F1) = H(v_1) - \frac{\left|v_{1_{F1=T}}\right|}{|v_1|}H\left(v_{1_{F1=T}}\right) - \frac{\left|v_{1_{F1=F}}\right|}{|v_1|}H\left(v_{1_{F1=F}}\right) = 1 - \frac{4}{8}*1 - \frac{4}{8}*1 = 0$$

$$IG(v_1, F2) = H(v_1) - \frac{\left|v_{1_{F2=T}}\right|}{|v_1|}H\left(v_{1_{F2=T}}\right) - \frac{\left|v_{1_{F2=F}}\right|}{|v_1|}H\left(v_{1_{F2=F}}\right) = 1 - \frac{2}{8}*0 - \frac{6}{8}* \left(-\frac{2}{3}(1-\log_2 3)\right) = 1 + \frac{1}{2}(1-\log_2 3) = \frac{3}{2} - \frac{1}{2}\log_2 3 \geq \frac{3}{2} - \frac{1}{2}\log_2 4 = \frac{3}{2} - \frac{1}{2}*2 = \frac{1}{2} > 0$$

$$IG(v_1, F3) = H(v_1) - \frac{\left|v_{1_{F3=T}}\right|}{|v_1|}H\left(v_{1_{F3=T}}\right) - \frac{\left|v_{1_{F3=F}}\right|}{|v_1|}H\left(v_{1_{F3=F}}\right) = 1 - \frac{4}{8}*1 - \frac{4}{8}*1 = 0$$

$$IG(v_1, F4) = H(v_1) - \frac{\left|v_{1_{F4=T}}\right|}{|v_1|}H\left(v_{1_{F4=T}}\right) - \frac{\left|v_{1_{F4=F}}\right|}{|v_1|}H\left(v_{1_{F4=F}}\right) = 1 - \frac{4}{8}*1 - \frac{4}{8}*1 = 0$$

Therefore, the split of the root will be according to the feature that maximizes the IG, and therefore it will be according to feature $F2$.

v_2-

$$H(v_2) = -\frac{4}{6}\log_2\frac{4}{6} - \frac{2}{6}\log_2\frac{2}{6} \cong 0.918$$

$$H\left(v_{2_{F1=T}}\right) = -\frac{2}{2}\log_2\frac{2}{2} - \frac{0}{2}\log_2 0 = 0$$

$$H\left(v_{2_{F1=F}}\right) = -\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4} = -\log_2\frac{2}{4} = -(1-2) = 1$$

$$H\left(v_{2_{F2=F}}\right) = -\frac{4}{6}\log_2\frac{4}{6} - \frac{2}{6}\log_2\frac{2}{6} \cong 0.918$$

$$H\left(v_{2_{F3=T}}\right) = -\frac{2}{4}\log_2\frac{2}{4} - \frac{2}{4}\log_2\frac{2}{4} = -\log_2\frac{2}{4} = 1$$

$$H\left(v_{2_{F3=F}}\right) = -\frac{2}{2}\log_2\frac{2}{2} - \frac{0}{2}\log_2\frac{0}{2} = 0$$

$$H\left(v_{2_{F4=T}}\right) = H\left(v_{2_{F4=F}}\right) = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} \cong 0.918$$

$$IG(v_2, F1) = H(v_2) - \frac{\left|v_{2_{F1=T}}\right|}{|v_2|}H\left(v_{2_{F1=T}}\right) - \frac{\left|v_{2_{F1=F}}\right|}{|v_2|}H\left(v_{2_{F1=F}}\right) = 0.918 - \frac{2}{6}*0 - \frac{4}{6}*1 = 0.251$$

$$IG(v_2, F2) = H(v_2) - \frac{\left|v_{2_{F2=T}}\right|}{|v_2|}H\left(v_{2_{F2=T}}\right) - \frac{\left|v_{2_{F2=F}}\right|}{|v_2|}H\left(v_{2_{F2=F}}\right) = 0.918 - 0 - \frac{6}{6}*0.918 = 0$$

$$IG(v_2, F3) = H(v_2) - \frac{\left|v_{2_{F3=T}}\right|}{|v_2|}H\left(v_{2_{F3=T}}\right) - \frac{\left|v_{2_{F3=F}}\right|}{|v_2|}H\left(v_{2_{F3=F}}\right) = 0.918 - \frac{4}{6}*1 - \frac{2}{6}*0 = 0.251$$

$$IG(v_2, F4) = H(v_2) - \frac{|v_{2_{F4=T}}|}{|v_2|} H(v_{2_{F4=T}}) - \frac{|v_{2_{F4=F}}|}{|v_2|} H(v_{2_{F4=F}}) = 0.918 - \frac{3}{6} * 0.918 - \frac{3}{6} * 0.251 = 0.3335$$

Therefore, the split of the root will be according to the feature that maximizes the IG, and therefore it will be according to feature $F4$.

$v_3 - Red\ Leaf$

$v_4 -$

$$H(v_4) = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} \cong 0.918$$

$$H(v_{4_{F1=T}}) = H(v_{4_{F3=F}}) = -\frac{1}{1}\log_2\frac{1}{1} - \frac{0}{1}\log_2\frac{0}{1} = 0$$

$$H(v_{4_{F1=F}}) = H(v_{4_{F3=T}}) = -\frac{1}{2}\log_2\frac{1}{2} - -\frac{1}{2}\log_2\frac{1}{2} = 1$$

$$H(v_{4_{F2=F}}) = H(v_{4_{F4=T}}) = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} \cong 0.918$$

$$IG(v_4, F1) = H(v_4) - \frac{|v_{4_{F1=T}}|}{|v_4|} H(v_{4_{F1=T}}) - \frac{|v_{4_{F1=F}}|}{|v_4|} H(v_{4_{F1=F}}) = 0.918 - \frac{1}{3} * 0 - \frac{2}{3} * 1 = 0.251$$

$$IG(v_4, F2) = H(v_4) - \frac{|v_{4_{F2=T}}|}{|v_4|} H(v_{4_{F2=T}}) - \frac{|v_{4_{F2=F}}|}{|v_4|} H(v_{4_{F2=F}}) = 0.918 - 0 - \frac{3}{3} * 0.918 = 0$$

$$IG(v_4, F3) = H(v_4) - \frac{|v_{4_{F3=T}}|}{|v_4|} H(v_{4_{F3=T}}) - \frac{|v_{4_{F3=F}}|}{|v_4|} H(v_{4_{F3=F}}) = 0.918 - \frac{2}{3} * 1 - \frac{1}{3} * 0 = 0.251$$

$$IG(v_4, F4) = H(v_4) - \frac{|v_{4_{F4=T}}|}{|v_4|} H(v_{4_{F4=T}}) - \frac{|v_{4_{F4=F}}|}{|v_4|} H(v_{4_{F4=F}}) = 0.918 - \frac{3}{3} * 0.918 - 0 = 0$$

Therefore, the split of the root will be according to the feature that maximizes the IG, and therefore it will be according to feature $F1$ $(or\ F3)$.

$v_5 -$

$$H(v_5) = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} \cong 0.918$$

$$H(v_{5_{F1=T}}) = H(v_{5_{F3=F}}) = -\frac{1}{1}\log_2\frac{1}{1} - \frac{0}{1}\log_2\frac{0}{1} = 0$$

$$H(v_{5_{F1=F}}) = H(v_{5_{F3=T}}) = -\frac{1}{2}\log_2\frac{1}{2} - -\frac{1}{2}\log_2\frac{1}{2} = 1$$

$$H(v_{5_{F2=F}}) = H(v_{5_{F4=F}}) = -\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3} \cong 0.918$$

$$IG(v_5, F1) = H(v_5) - \frac{|v_{5_{F1=T}}|}{|v_5|} H(v_{5_{F1=T}}) - \frac{|v_{5_{F1=F}}|}{|v_5|} H(v_{5_{F1=F}}) = 0.918 - \frac{1}{3} * 0 - \frac{2}{3} * 1 = 0.251$$

$$IG(v_5, F2) = H(v_5) - \frac{|v_{5_{F2=T}}|}{|v_5|} H(v_{5_{F2=T}}) - \frac{|v_{4_{F2=F}}|}{|v_5|} H(v_{5_{F2=F}}) = 0.918 - 0 - \frac{3}{3} * 0.918 = 0$$

$$IG(v_5, F3) = H(v_5) - \frac{|v_{5_{F3=T}}|}{|v_5|} H(v_{5_{F3=T}}) - \frac{|v_{5_{F3=F}}|}{|v_5|} H(v_{5_{F3=F}}) = 0.918 - \frac{2}{3} * 1 - \frac{1}{3} * 0 = 0.251$$

$$IG(v_5, F4) = H(v_5) - \frac{|v_{4_{F4=T}}|}{|v_4|} H(v_{4_{F4=T}}) - \frac{|v_{4_{F4=F}}|}{|v_4|} H(v_{4_{F4=F}}) = 0.918 - 0 - \frac{3}{3} * 0.918 - 0 = 0$$

Therefore, the split of the root will be according to the feature that maximizes the IG, and therefore it will be according to feature $F1$ $(or\ F3)$.

$v_6 - Green\ Leaf$

$v_7 -$

$$H(v_7) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$$

$$H\left({v_7}_{F1=F}\right) = H\left({v_7}_{F2=F}\right) = H\left({v_7}_{F4=T}\right) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$$

$$H\left({v_7}_{F3=F}\right) = H\left({v_7}_{F3=T}\right) = -1\log_2 1 - 0 = 0$$

$$IG(v_7, F1) = H(v_7) - \frac{\left|{v_7}_{F1=T}\right|}{|v_7|}H\left({v_7}_{F1=T}\right) - \frac{\left|{v_7}_{F1=F}\right|}{|v_7|}H\left({v_7}_{F1=F}\right) = 1 - 0 - 1 = 0$$

$$IG(v_7, F2) = H(v_7) - \frac{\left|{v_7}_{F2=T}\right|}{|v_7|}H\left({v_7}_{F2=T}\right) - \frac{\left|{v_7}_{F2=F}\right|}{|v_7|}H\left({v_7}_{F2=F}\right) = 1 - 0 - 1 = 0$$

$$IG(v_7, F3) = H(v_7) - \frac{\left|{v_7}_{F3=T}\right|}{|v_7|}H\left({v_7}_{F3=T}\right) - \frac{\left|{v_7}_{F3=F}\right|}{|v_7|}H\left({v_7}_{F3=F}\right) = 1 - \frac{1}{2} * 0 - \frac{1}{2} * 0 = 1$$

$$IG(v_7, F4) = H(v_7) - \frac{\left|{v_7}_{F4=T}\right|}{|v_7|}H\left({v_7}_{F4=T}\right) - \frac{\left|{v_7}_{F4=F}\right|}{|v_7|}H\left({v_7}_{F4=F}\right) = 1 - \frac{2}{2} * 1 - 0 = 0$$

Therefore, the split of the root will be according to the feature that maximizes the IG, and therefore it will be according to feature $F3$.

$v_8 - Green\ Leaf$

$v_9 -$

$$H(v_9) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$$

$$H\left({v_9}_{F1=F}\right) = H\left({v_9}_{F2=F}\right) = H\left({v_9}_{F4=F}\right) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$$

$$H\left({v_9}_{F3=F}\right) = H\left({v_9}_{F3=T}\right) = -1\log_2 1 - 0 = 0$$

$$IG(v_9, F1) = H(v_9) - \frac{\left|{v_9}_{F1=T}\right|}{|v_9|}H\left({v_9}_{F1=T}\right) - \frac{\left|{v_9}_{F1=F}\right|}{|v_9|}H\left({v_9}_{F1=F}\right) = 1 - 0 - 1 = 0$$

$$IG(v_9, F2) = H(v_7) - \frac{\left|{v_9}_{F2=T}\right|}{|v_9|}H\left({v_9}_{F2=T}\right) - \frac{\left|{v_9}_{F2=F}\right|}{|v_9|}H\left({v_9}_{F2=F}\right) = 1 - 0 - 1 = 0$$

$$IG(v_9, F3) = H(v_9) - \frac{\left|{v_9}_{F3=T}\right|}{|v_9|}H\left({v_9}_{F3=T}\right) - \frac{\left|{v_9}_{F3=F}\right|}{|v_9|}H\left({v_9}_{F3=F}\right) = 1 - \frac{1}{2} * 0 - \frac{1}{2} * 0 = 1$$

$$IG(v_9, F4) = H(v_9) - \frac{\left|{v_9}_{F4=T}\right|}{|v_9|}H\left({v_9}_{F4=T}\right) - \frac{\left|{v_9}_{F4=F}\right|}{|v_9|}H\left({v_9}_{F4=F}\right) = 1 - 0 - \frac{2}{2} * 1 = 0$$

Therefore, the split of the root will be according to the feature that maximizes the IG, and therefore it will be according to feature $F3$.

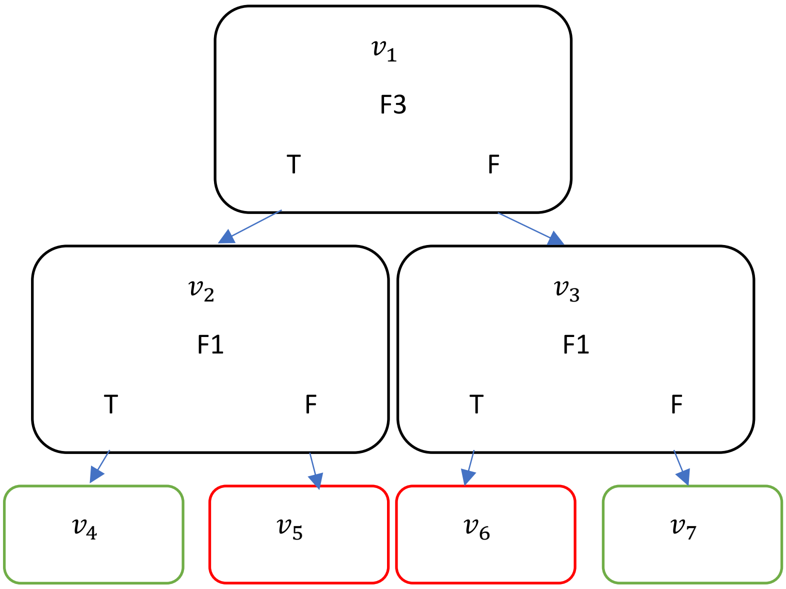
$v_{10} - Red\ Leaf$

$v_{11} - Green\ Leaf$

$v_{12} - Read\ Leaf$

$v_{13} - Green\ Leaf$

There exists a perfect tree of depth 2:



If we will limit ID3 to max depth of 2, after the first split of the root, which be exactly as in the above ID3 tree (F2 feature), two data entries will go to the right node and be classified as False, and the other six data entries will go to the left. Afterwards the node will be split, as it happened in the above ID3 tree, according to the F4 feature, as a result three data entries will go left and three data entries will go right, and at each of them one entry is supposed to be classified as False, and the other two as True, and therefore, because we reached the maximum depth of 2, we will classify according to the majority, and therefore both nodes will be classified as True (we can actually make the parent of those two nodes as True and save space).

Therefore, for each of the two data entries in the test dataset which were False, but we classified them as True because of the limitation of the depth, we will get an error. For all the other six samples we will return the correct answer, and therefore the empirical error is as follows:

$$Empirical\ Error = \frac{|Wrongly\ \ classified\ samples\ of\ the\ test\ dataset|}{|Test\ Dataset|} = \frac{2}{8} = \frac{1}{4}$$

Q2

The statement is False.

For example, given a tree node with the following data entries:

#	$\boldsymbol{a}$	**LABEL**
1	T	T
2	T	F
3	F	F
4	F	F
5	F	F

The Entropy of the node is:

$$H(v) = -\frac{1}{5} * \log_2 \frac{1}{5} - \frac{4}{5} \log_2 \frac{4}{5}$$

After we split using feature a we will receive the following entropies,

$$H(v_{a=T}) = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1$$

$$H(v_{a=F}) = -0 - \frac{4}{4} \log_2 1 = 0$$

It holds that $H(v) \cong 0.722 < 1 = H(v_{a=T})$, and therefore it is a counterexample.

Q3

1

MODEL\DATASET	(A)	(B)	(C)
I	Yes.	No. There are two very close opposite classification data points at $(\sim 3, \pm\epsilon)$, and therefore their classification would be wrong, because each of them will receive the classification of the other point.	No. Each pair of dots that has the same y axis, are of the opposite class, and they are the closest neighbor of the other dot, and therefore they will make the other dot to be classified as the opposite class.
II	No. Because for each point, the 3 (closest) neighbors it has contain 2 neighbors of the opposite classification, and therefore the KNN classification will be the opposite of the required one.	Yes.	No. For example for the upper left dot(which is blue) the closest 3 neighbors are 2 of class red and one that of type blue, and therefore it will be (wrongfully) classified as red.
III	Yes.	No. Because each line that will separate the data will have samples of the two classified classes on each side and therefore will result in erroneous classifications.	Yes.
IV	Yes.	Yes.	Yes.

2

i. Unchanged, because distance of each pair of points will remain the same after the rotation, and therefore the closest neighbors will remain the closest neighbors.

ii. Unchanged, because distance of each pair of points will remain the same after the rotation, and therefore the closest neighbors will remain the closest neighbors.

iii. Unchanged, because for datasets it could solve previously, each homogeneous linear model that could solve a previous dataset, can be rotated by the same angel and it will make the exact same decisions as previously. In practice, the modified homogeneous linear model can rotate the points back to their original coordinates, and afterwards to do the same comparison as the original model did. For datasets it could not solve previously, if it could solve them after the rotation, we could solve for the original dataset with training error of 0 by rotating the dataset to the new angle, in conjunction for the fact it could not be solved with training error of 0.

iv. The answers may change to dataset (B) and (C).

For dataset (A), we need to do at most 2 splits to properly split the two sides of classes, be splitting according to the coordinates of an imaginary point in between both sides of the data. Because this (new) point coordinates are in $\mathbb{R}^2$, we could get the required empirical error by first split according to the x-axis and for each child node a split by y-axis which will result in 4 leaves.

For dataset (B) the answer may change because after the rotation the regions of data may not be easily separable by two splits as done in dataset (A). Each time we need to split the data according to both axes, we will result at least in 4 new leaves each time. In this dataset, we need to split to more than 2 regions according to both axis and therefore will result in more than 4 leaves.

For dataset (C), like dataset (B), after the rotation, the data may not be easily separable and we will have to do 2 comparisons for x and y axis, for each dot, which will result in more than 4 leaves.

For clarification purposes, whenever there is a need to split the data according to both axes, (i.e., splitting according to only a single axis will not be sufficient), we must do two comparisons (for example, as done in (A) dataset), one for each of the axis to perfectly split the data, which will result in more new leaves.