

# Machine Learning

## Assignment 5.2

Submitted By: Ranji Raj

November 27, 2020

### Rules for Pruning

- If the error on the tree after pruning is **same** as the previous/original, then retain the latest tree.
- If the error on the tree after pruning is **more** as the previous, then retain the previous tree.
- If the error on the tree after pruning is **less** as the previous, then accept the new tree and discard the previous tree.

### Calculating Error for the first time

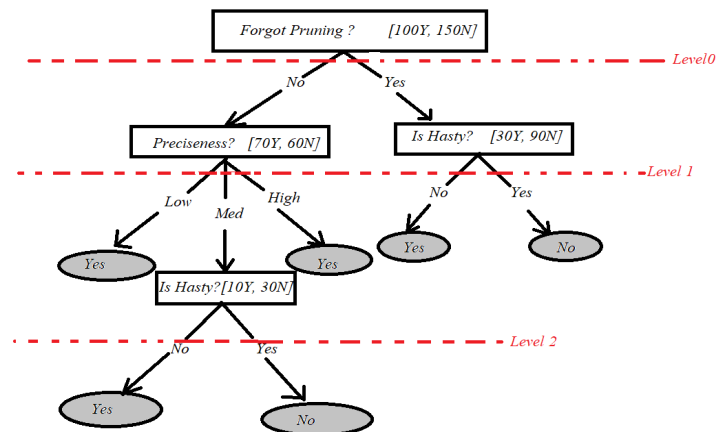


Figure 1: Original Tree

id	Forgot Pruning	Preciseness	Is Hasty	Full Points
1	yes	low	no	no
2	yes	high	yes	no
3	no	low	no	yes
4	no	med	no	no
5	no	high	yes	yes
6	yes	med	no	no

Table 1: Validation set

id	Forgot Pruning	Preciseness	Is Hasty	Full Points	Is Correct
1	yes	low	no	no	False
2	yes	high	yes	no	True
3	no	low	no	yes	True
4	no	med	no	no	False
5	no	high	yes	yes	True
6	yes	med	no	no	False

Table 2: Verification check 1

Calculating the error on the complete tree as,

$$ErrorRate, e = \frac{FP + FN}{TP + FP + FN + TN}$$

$$\therefore e_1 = \frac{3}{6} = \mathbf{0.5}$$

## First Prune

The error rate is very high, so, we prune the left sub-tree at level 2 and recalculate the error, We look at the class distribution at node *Is Hasty?*, the majority class here is *No* hence, we replace it by a leaf.

id	Forgot Pruning	Preciseness	Is Hasty	Full Points	Is Correct
1	yes	low	no	no	False
2	yes	high	yes	no	True
3	no	low	no	yes	True
4	no	med	no	no	True
5	no	high	yes	yes	True
6	yes	med	no	no	False

Table 3: Verification check 2

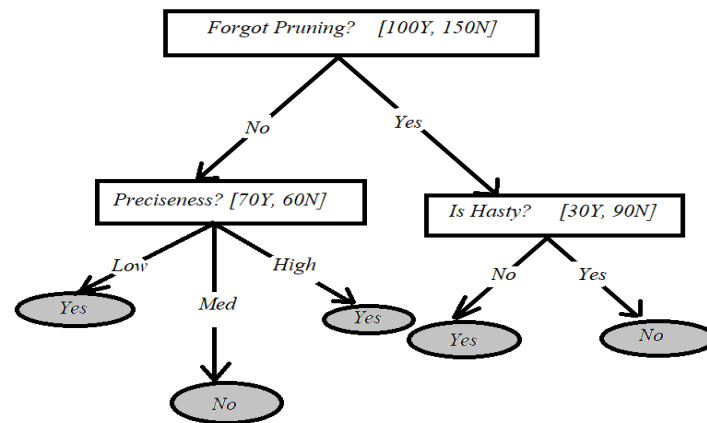


Figure 2: First Pruned Tree

$$e_2 = \frac{2}{6} = \mathbf{0.33}$$

We observe the error has reduced significantly, so retain the first pruned tree.

## Second Prune

We shall now prune the right sub-tree at level 1. We look at the class distribution at node *Is Hasty?*, the majority class here is *No* hence, we replace it by a leaf.

id	Forgot Pruning	Preciseness	Is Hasty	Full Points	Is Correct
1	yes	low	no	no	True
2	yes	high	yes	no	True
3	no	low	no	yes	True
4	no	med	no	no	True
5	no	high	yes	yes	True
6	yes	med	no	no	True

Table 4: Verification check 3

$$e_3 = \frac{0}{6} = \mathbf{0.0}$$

We observe no error, so accept the second pruned tree.

## Third Prune

We shall now prune the left sub-tree at level 1 and recalculate the error, we look at the class distribution at node *Preciseness?*, the majority class here is *Yes* hence, we replace

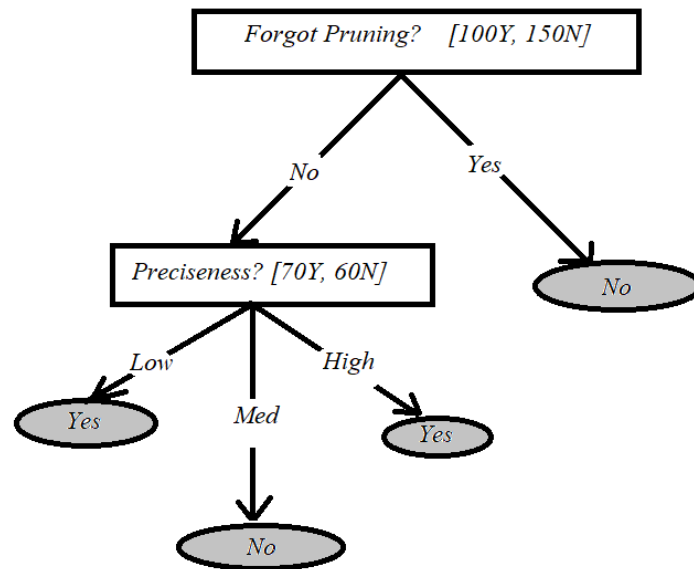


Figure 3: Second Pruned Tree

it by a leaf.

id	Forgot Pruning	Preciseness	Is Hasty	Full Points	Is Correct
1	yes	low	no	no	True
2	yes	high	yes	no	True
3	no	low	no	yes	True
4	no	med	no	no	False
5	no	high	yes	yes	True
6	yes	med	no	no	True

Table 5: Verification check 4

$$e_4 = \frac{1}{6} = \mathbf{0.1667}$$

We observe a rise in the error rate, so reject the third pruned tree and keep the second pruned tree.

## Fourth Prune

We shall now prune the root at level 0 and convert into a leaf node, we look at the class distribution at node *Forgot Pruning?*, the majority class here is *No* hence, we replace it by a leaf.

$$e_5 = \frac{2}{6} = \mathbf{0.333}$$

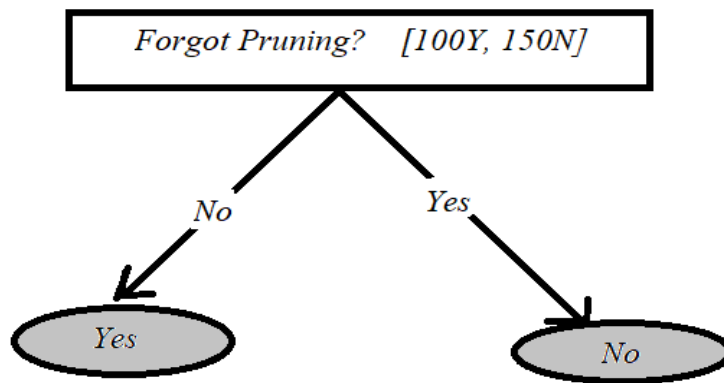


Figure 4: Third Pruned Tree

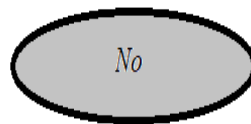


Figure 5: Fourth Pruned Tree

We observe a rise in the error rate again, and moreover, it makes the tree look biased. So reject the fourth pruned tree and accept the second pruned tree as the final tree.

id	Forgot Pruning	Preciseness	Is Hasty	Full Points	Is Correct
1	yes	low	no	no	True
2	yes	high	yes	no	True
3	no	low	no	yes	False
4	no	med	no	no	True
5	no	high	yes	yes	False
6	yes	med	no	no	True

Table 6: Verification check 5

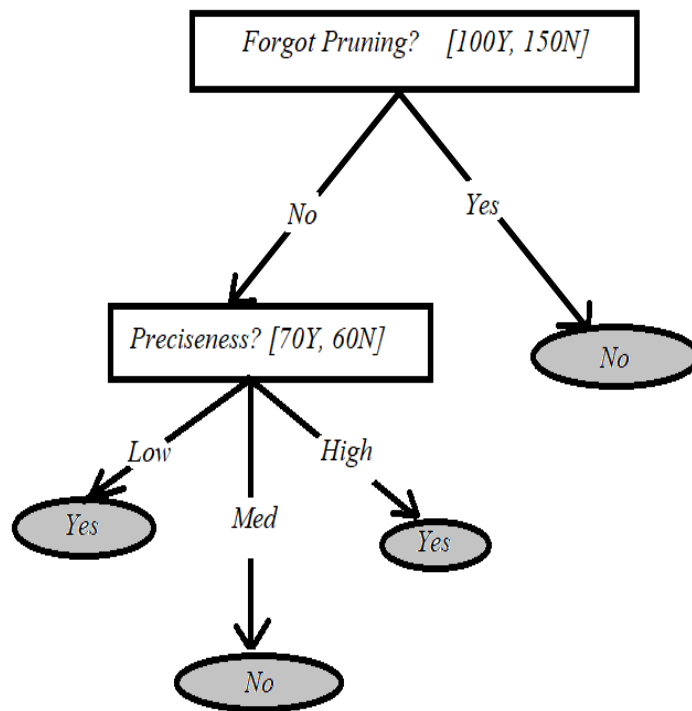


Figure 6: Final Tree