7. The following table summarizes a data set with three attributes A, B, C and two class labels $+$, $-$. Build a two-level decision tree.

A	B	C	Number of Instances	
			+	−
T	T	T	5	0
F	T	T	0	20
T	F	T	20	0
F	F	T	0	5
T	T	F	0	0
F	T	F	25	0
T	F	F	0	0
F	F	F	0	25

(a) According to the classification error rate, which attribute would be chosen as the first splitting attribute? For each attribute, show the contingency table and the gains in classification error rate.

Answer:

The error rate for the data without partitioning on any attribute is

$$E_{orig} = 1 - \max(\frac{50}{100}, \frac{50}{100}) = \frac{50}{100}.$$

After splitting on attribute A, the gain in error rate is:

	$A = T$	$A = F$
+	25	25
−	0	50

$$E_{A=T} = 1 - \max(\frac{25}{25}, \frac{0}{25}) = \frac{0}{25} = 0$$

$$E_{A=F} = 1 - \max(\frac{25}{75}, \frac{50}{75}) = \frac{25}{75}$$

$$\Delta_A = E_{orig} - \frac{25}{100}E_{A=T} - \frac{75}{100}E_{A=F} = \frac{25}{100}$$

After splitting on attribute B, the gain in error rate is:

	$B = T$	$B = F$
+	30	20
−	20	30

$$E_{B=T} = \frac{20}{50}$$

$$E_{B=F} = \frac{20}{50}$$

$$\Delta_B = E_{orig} - \frac{50}{100}E_{B=T} - \frac{50}{100}E_{B=F} = \frac{10}{100}$$

After splitting on attribute C, the gain in error rate is:

	$C=T$	$C=F$
+	25	25
−	25	25

$$E_{C=T} = \frac{25}{50}$$
$$E_{C=F} = \frac{25}{50}$$
$$\Delta_C = E_{orig} - \frac{50}{100}E_{C=T} - \frac{50}{100}E_{C=F} = \frac{0}{100} = 0$$

The algorithm chooses attribute A because it has the highest gain.

(b) Repeat for the two children of the root node.

Answer:

Because the $A=T$ child node is pure, no further splitting is needed. For the $A=F$ child node, the distribution of training instances is:

B	C	Class label	
		+	−
T	T	0	20
F	T	0	5
T	F	25	0
F	F	0	25

The classification error of the $A=F$ child node is:

$$E_{orig} = \frac{25}{75}$$

After splitting on attribute B, the gain in error rate is:

	$B=T$	$B=F$
+	25	0
−	20	30

$$E_{B=T} = \frac{20}{45}$$
$$E_{B=F} = 0$$
$$\Delta_B = E_{orig} - \frac{45}{75}E_{B=T} - \frac{20}{75}E_{B=F} = \frac{5}{75}$$

After splitting on attribute C, the gain in error rate is:

	$C=T$	$C=F$
+	0	25
−	25	25

$$E_{C=T} = \frac{0}{25}$$
$$E_{C=F} = \frac{25}{50}$$
$$\Delta_C = E_{orig} - \frac{25}{75}E_{C=T} - \frac{50}{75}E_{C=F} = 0$$

The split will be made on attribute B.

(c) How many instances are misclassified by the resulting decision tree

Answer:

20 instances are misclassified. (The error rate is $\frac{20}{100}$.)

(d) Repeat parts (a), (b), and (c) using C as the splitting attribute.

Answer:

For the $C = T$ child node, the error rate before splitting is:
$E_{orig} = \frac{25}{50}$.

After splitting on attribute A, the gain in error rate is:

	$A = T$	$A = F$
+	25	0
−	0	25

$$E_{A=T} = 0$$
$$E_{A=F} = 0$$
$$\Delta_A = \frac{25}{50}$$

After splitting on attribute B, the gain in error rate is:

	$B = T$	$B = F$
+	5	20
−	20	5

$$E_{B=T} = \frac{5}{25}$$
$$E_{B=F} = \frac{5}{25}$$
$$\Delta_B = \frac{15}{50}$$

Therefore, A is chosen as the splitting attribute.

For the $C = F$ child, the error rate before splitting is: $E_{orig} = \frac{25}{50}$.

After splitting on attribute A, the error rate is:

	$A = T$	$A = F$
+	0	25
−	0	25

$$E_{A=T} = 0$$
$$E_{A=F} = \frac{25}{50}$$
$$\Delta_A = 0$$

After splitting on attribute B, the error rate is:

	$B = T$	$B = F$
+	25	0
−	0	25

$$E_{B=T} = 0$$
$$E_{B=F} = 0$$
$$\Delta_B = \frac{25}{50}$$

Therefore, B is used as the splitting attribute.

The overall error rate of the induced tree is 0.

(e) Use the results in parts (c) and (d) to conclude about the greedy nature of the decision tree induction algorithm.

The greedy heuristic does not necessarily lead to the best tree.