- Problem-2: Decision Trees based on GINI Index

(1)

A	B	Class Label	
		+	−
T	T	0	20
T	F	20	10
F	T	15	0
F	F	0	35

	A=T	A=F	B=T	B=F	Total
+	20	15	15	20	70
−	30	35	90	45	130

⇒ Now, computing GINI,

$$GINI(t) = 1 - \sum_j [P(j|t)]^2$$

$$= 1 - \left(\frac{70}{200}\right)^2 - \left(\frac{130}{200}\right)^2$$

$$= 1 - (0.35)^2 - (0.65)^2$$

$$\boxed{GINI(t) = 0.455}$$

⇒ Now, computing $GINI(A=T) = 1 - \left(\left(\frac{20}{50}\right)^2 + \left(\frac{30}{50}\right)^2\right)$

$$= 1 - \left((0.4)^2 + (0.6)^2\right)$$

$$= 0.48$$

⇒ Now, computing $GINI(A=F) = 1 - \left(\frac{15}{50}\right)^2 - \left(\frac{35}{50}\right)^2$

$$= \underline{0.42}$$

$$GINI(A) = \frac{50}{100}(0.48) + \frac{50}{100}(0.42)$$

$$= 0.24 + 0.21$$

$$= \underline{0.45}$$

⇒ Now, computing $GINI(B=T) = 1 - \left[\left(\frac{15}{35}\right)^2 + \left(\frac{20}{35}\right)^2\right]$

$= 0.4897$

⇒ Now, computing $GINI(B=F) = 1 - \left[\left(\frac{20}{65}\right)^2 + \left(\frac{45}{65}\right)^2\right]$

$= 0.4260$

⇒ $GINI(B) = \frac{35}{100} \times 0.4897 + \frac{65}{100} \times 0.4260$

$= \underline{0.4482}$

Therefore, $GINI(B) < GINI(A)$

Hence, Attribute B is the ~~best~~ split.

(2)

Cost Matrix	Attribute Value		
		T	F
Actual class	+	-1	100
	−	0	-10

⇒ Cost matrix for A,

$= (20 \times -1) + (30 \times 0) + (15 \times 100) + (35 \times -10)$

$= -20 + 0 + 1500 - 350$

$= \underline{1130}$

⇒ Cost matrix for B,

$= (15 \times -1) + (20 \times 0) + (20 \times 100) + (45 \times -10)$

$= -15 + 0 + 2000 - 450$

$= \underline{1535}$

∴ The cost error for A is less than cost error for B.
So, we can use A for the first splitting attribute.