Q2.

**(a).**

H(Target Class) = -Σp(Target Class)ilog2p(Target Class)i

P(Class=`+`) = 4/9

P(Class=`-`) = 5/9

H(Target Class) = -4/9log24/9 – 5/9log25/9

= -4/9(-1.169) – 5/9(-0.848)

= 0.5199 + 0.471

H(Target Class) = 0.991

**(b).**

H(Target Class/a1) = 4/9(-3/4log23/4 – 1/4log21/4) + 5/9(-1/5log21/5 – 4/5log24/5)

= 0.36 + 0.401

= 0.761

H(Target Class/a2) = 5/9(-2/5log22/5 – 3/5log23/5) + 4/9(-2/4log22/4 – 2/4log22/4)

= 0.539 + 0.444

= 0.983

I.G.(Target/a1) = H(Target Class) - H(Target Class/a1)

= 0.991 – 0.761

= 0.23

I.G.(Target/a2) = H(Target Class) - H(Target Class/a2)

= 0.991 – 0.983

= 0.008

(c).

Table in increasing order of a3 values is:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| A3 | 1.0 | 3.0 | 4.0 | 5.0 | 5.0 | 6.0 | 7.0 | 7.0 | 8.0 |
| Target Class | + | - | + | - | - | + | - | + | - |

I.G. (a3=1.0) = H(Target Class) - H(Target Class/a3=1.0)

= 0.991 – [1(-4/9log24/9 – 5/9log25/9)]

= 0

I.G. (a3=3.0) = H(Target Class) - H(Target Class/a3=3.0)

= 0.991 – [1/9(-1log21) + 8/9(-3/8log23/8 – 5/8log25/8)]

= 0.991 – (0 + 0.848)

= 0.143

I.G. (a3=4.0) = H(Target Class) - H(Target Class/a3=4.0)

= 0.991 – [2/9(-1/2log21/2 – 1/2log21/2)+ 7/9(-3/7log23/7 – 4/7log24/7)]

= 0.991 – (0.222 + 0.766)

= 0.003

I.G. (a3=5.0) = H(Target Class) - H(Target Class/a3=5.0)

= 0.991 – [3/9(-2/3log22/3 – 1/3log21/3)+ 6/9(-2/6log22/6 – 4/6log24/6)]

= 0.991 – (0.306 + 0.612)

= 0.073

I.G. (a3=6.0) = H(Target Class) - H(Target Class/a3=6.0)

= 0.991 – [5/9(-2/5log22/5 – 3/5log23/5)+ 4/9(-2/4log22/4 – 2/4log22/4)]

= 0.991 – (0.539 + 0.444)

= 0.008

I.G. (a3=7.0) = H(Target Class) - H(Target Class/a3=7.0)

= 0.991 – [6/9(-3/6log23/6 – 3/6log23/6)+ 3/9(-1/3log21/3 – 2/3log22/3)]

= 0.991 – (0.667 + 0.306)

= 0.018

I.G. (a3=8.0) = H(Target Class) - H(Target Class/a3=8.0)

= 0.991 – [8/9(-4/8log24/8 – 4/8log24/8)+ 1/9(0 – 1log21)]

= 0.991 – (0.889 + 0)

= 0.102

As, Maximum Information Gain is from a3=3.0, choose a3=3.0 as best split point attribute a3, that is a3<3 and a3≥3.

(d).

Information Gain (I.G.) for a1, a2 and a3 are as follows:

I.G. (a1) = 0.23

I.G. (a2) = 0.008

I.G. (a3) = 0.143

As, maximum Information Gain is from attribute a1, algorithm will choose a1 as root of the tree.

(e).

GINI (t) = 1 - Σj [p( j|t )]2

GINI ( Target Class) = 1 – (4/9)2 – (5/9)2

= 0.494

GINI (a1, T) = 1 – (3/4)2 – (1/4)2

= 0.375

GINI (a1, F) = 1 – (1/5)2 – (4/5)2

= 0.32

GINIsplit = Σi=1to k  ni/n GINI (i)

GINI (a1) = 4/9(0.375) + 5/9(0.32)

= 0.344

GINI (a2, T) = 1 – (2/5)2 – (3/5)2

= 0.48

GINI (a2, F) = 1 – (2/4)2 – (2/4)2

= 0.5

GINI (a2) = 5/9(0.48) + 4/9(0.5)

= 0.489

According to GINI Index, a1 is the better attribute to split (between a1 and a2) as GINI Index is smallest for a1.

(f).

Error ( t ) = 1 – Max P( i | t )

Error (Target Class) = 1 – Max( 4/9, 5/9 )

= 1 – 5/9

= 4/9

Error (a1, T) = 1 – Max ( 3/4, 1/4 )

= 1 – 3/4

= 1/4

Error (a1, F) = 1 – Max ( 1/5, 4/5 )

= 1 – 4/5

= 1/5

Error (a1) = P(T)(Error (a1, T)) + P(F)(Error (a1, F))

= 4/9(1/4) + 5/9(1/5)

= 2/9

Error (a2, T) = 1 – Max ( 2/5, 3/5 )

= 1 – 3/5

= 2/5

Error (a2, F) = 1 – Max ( 2/4, 2/4 )

= 1 – 2/4

= 2/4

Error (a2) = P(T)(Error (a2, T)) + P(F)(Error (a2, F))

= 5/9(2/5) + 4/9(2/4)

= 4/9

As, Misclassification Error of a1, Error (a1), is minimum, So, Information Gain of a1 is maximum. Therefore, a1 is better attribute according to misclassification error.

**Decision Tree:**

As, a1 is the first attribute we have chosen, according to part (d), to be root. Let’s construct it and observe the remaining elements based on choosing the best feature to split.

**For a1 = T :**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **a2** | T | F | T | F |
| **a3** | 1.0 | 5.0 | 6.0 | 7.0 |
| **Target Class** | + | - | + | + |

a2 a3: Clearly there is no point to choose for splitting

+ : 2

- : 0

+ : 1

- : 1

T

F

Ab. Sure Ab. Unsure

So, we will choose a2 for splitting.

**For a1 = F :**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **a2** | T | F | T | T | F |
| **a3** | 3.0 | 4.0 | 5.0 | 7.0 | 8.0 |
| **Target Class** | - | + | - | - | - |

a2 a3

+ : 1

- : 1

+ : 0

- : 3

<5

≥5

+ : 0

- : 3

+ : 1

- : 1

T

F

Ab. Sure Ab. Unsure Ab. Unsure Ab. Sure

As both are same in terms of predictability, so, we can choose any of two for splitting. Let’s choose a2 for splitting.

Q3.

(a).

For Generalization Error Rate let’s observe the training dataset using the decision tree.

A=0 A=1

B=0 B=1 C=0 C=1

2 `+` 1 `+`, 2 `-` 2 `+`, 3 `-` 1 `+`

Here, number of errors are = 0 + 1 + 2 + 1

= 3

Generalization Error Rate = (No. of Errors / Total Number of Values) \* 100

= (3 / 10) \* 100

= 0.3 \* 100

= 30%

(b).

Confusion Matrix:

|  |  |  |
| --- | --- | --- |
|  | **T** | **F** |
| **T** | TP = 3 | FN = 0 |
| **F** | FP = 1 | TN = 1 |

Accuracy = TP+TN / TP+TN+FP+FN

= 4 / 5

= 80%

Precision = TP / TP+FP

= 3 / 4

= 75%

Recall = TP / TP+FN

= 3 / 3

= 1

= 100%