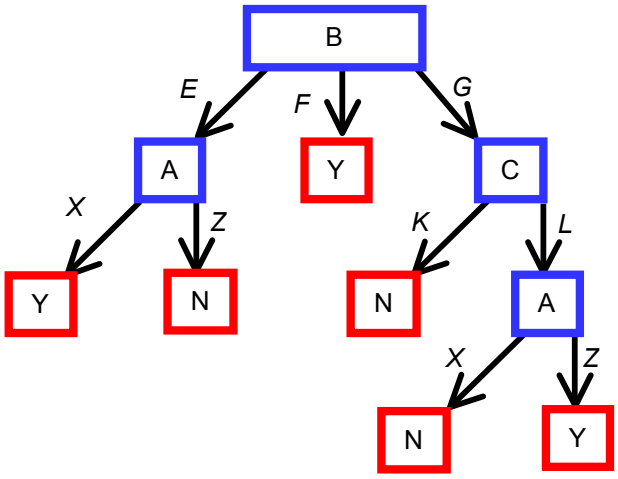
INTRODUCTION TO ARTIFICIAL INTELLIGENCE – LECTURE 3 – DECISION TREES

Name and surname: Sofya Aksenyuk Index no.: 150284

I. Given the below decision tree referring to condition attributes A, B, and C and a decision attribute D (class Y or N), answer the questions given below.



a) Compute the number of levels. **Answer**: 3.

b) Compute the number of leaves. **Answer**: 6

c) Compute the tree depth. **Answer**: 4.

d) Formulate the underlying decision rules for class N:

Answer: ... if (B = E and A = Z) then D = N
if (B = G and C = K) then D = N
if (B = G and C = L and A = X) then D = N

e) Formulate the disjunctions (∨ = OR) of conjunctions (∧ = AND) for class Y:

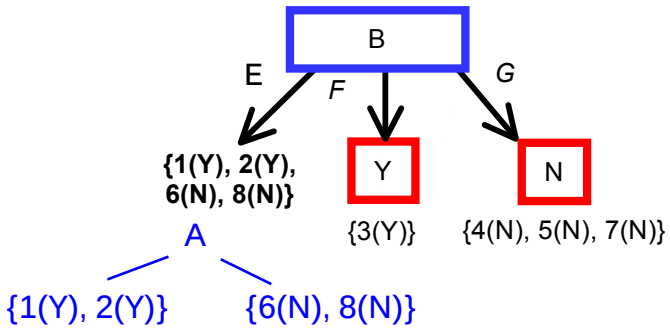
Answer: ... X(Y) = (B = F) ⋁ (B = E ⋀ A = X) ⋁ (B = G ⋀ C = L ⋀ A = Z)

f) Use the decision tree to classify (i.e., determine the value of decision attribute D) the two examples 1 and 2 with the descriptions provided in the below table:

	A	B	C	D
1	Z	E	L	N
2	Z	G	L	Y

II. Consider the below information table involving three condition attributes A, B, and C, and a single decision attribute D (class Y or N).

	A	B	C	D
1	X	E	K	Y
2	X	E	L	Y
3	X	F	L	Y
4	X	G	K	N
5	X	G	L	N
6	Z	E	L	N
7	Z	G	L	N
8	Z	E	K	N



a) The decision tree after the top-level split using *InformationGain* is presented above. Consider examples 1, 2, 6 and 8 in the left-most leaf, and solve the following tasks:

i) Knowing that for the left-most leaf:

- $Ent(D) = -2/4 \cdot log_2(2/4) - 2/4 \cdot log_2(2/4) = 1$
- $Ent(D,C) = 2/4(-1/2 \cdot log_2(1/2) - 1/2 \cdot log_2(1/2)) + 2/4(-1/2 \cdot log_2(1/2) - 1/2 \cdot log_2(1/2)) = 0$

$-\frac{1}{2}\cdot(-1) - \frac{1}{2}\cdot(-1) = 1 \Rightarrow \frac{2}{4} + \frac{2}{4} = 1$

compute $InformationGain(D,C)$ = Ent(D) - Ent(D, C) = 1 - 1 = 0

ii) Compute $Ent(D,A)$ and $InformationGain(D,A)$:

- $Ent(D,A) = \frac{2}{4}Ent(D_x) + \frac{2}{4}Ent(D_2) = \frac{2}{4}\left(-\frac{2}{2}\log_2\frac{2}{2} - \frac{0}{2}\log_2\frac{0}{2}\right) + \frac{2}{4}\left(-\frac{2}{2}\log_2\frac{2}{2} - \frac{0}{2}\log_2\frac{0}{2}\right) = \frac{2}{4}\cdot 0 + \frac{2}{4}\cdot 0 = 0$
- $InformationGain(D,A)$ = 1 - 0 = 1

iii) Which attribute, A or C, would be selected for the split in the left-most node when using *InformationGain*?

Answer: A

b) Consider the top-level split (i.e., assume no node has been created) using *GainRatio* rather than *InformationGain*, and answer the following questions.

i) Knowing that for the top-level split:

- $Ent(D) = -3/8 \cdot log_2(3/8) - 5/8 \cdot log_2(5/8) = 0.955$
- $Ent(A) = -3/8 \cdot log_2(3/8) - 5/8 \cdot log_2(5/8) = 0.955$
- $Ent(B) = -4/8 \cdot log_2(4/8) - 1/8 \cdot log_2(1/8) - 3/8 \cdot log_2(3/8) = 1.406$
- $InformationGain(D,A) = 0.955 - 0.607 = 0.348$
- $InformationGain(D,B) = 0.955 - 0.500 = 0.455$
- $InformationGain(D,C) = 0.955 - 0.951 = 0.004$

compute:

- $GainRatio(D,A) = \frac{InfGain(D,A)}{InfSplit(D,A)} = \frac{0,348}{0,955} = 0,364$
- $GainRatio(D,B)$ = 0,455 / 1,405 = 0,324

ii) Compute:

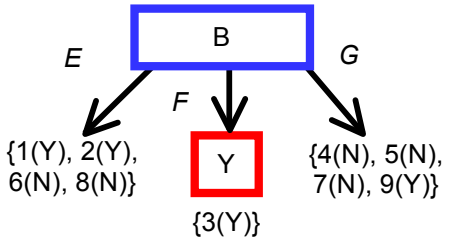
- $Ent(C) = -\frac{3}{8}\log_2\frac{3}{8} - \frac{5}{8}\log_2\frac{5}{8} = 0,531 + 0,424 = 0,955$
- $GainRatio(D,C)$ = 0,004 / 0,955 = 0,004

iii) Which attribute, A, B, or C, would be selected for the top-level split based on *GainRatio*?

Answer: A (0,364 > 0,324 > 0,004)

III. Given the below information table involving three condition attributes A, B, and C and a single decision attribute D (class Y or N), including additional example (9) implying the inconsistency, the below tree has been obtained after the first iteration using ID3 with *InformationGain*.

	A	B	C	D
1	X	E	K	Y
2	X	E	L	Y
3	X	F	L	Y
4	X	G	K	N
5	X	G	L	N
6	Z	E	L	N
7	Z	G	L	N
8	Z	E	K	N
9	Z	G	L	Y



Assuming that the tree is pre-pruned if the share of examples in a given node from a single class is at least 2/3, which node would be expanded and which not? What class would be assigned to the respective leaf/leaves)?

Answer:

- **left-most leaf**: expanded / not expanded;
  answer only if not expanded: the assigned class would be ...
- **right-most leaf**: expanded / not expanded;
  answer only if not expanded: the assigned class would be N