

## Домашна работа №2

### Задача 1.

Разгледайте следното множество от обучаващи примери: (T – истина; F – лъжа)

Пример	Класификация	A <sub>1</sub>	A <sub>2</sub>
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

- a) Каква е ентропията на това множество от обучаващи примери по отношение за целевата класификация?
- b) Каква е информационната печалба на атрибута A<sub>2</sub> по отношение на тези примери?

### Решение.

- a) Имаме 6 примера - 2 положителни (+) и 4 отрицателни (-)

$$\text{Entropy}(S) = -\frac{2}{6} \log_2 \frac{2}{6} - \frac{4}{6} \log_2 \frac{4}{6} \approx 0,918296$$

- b) Разделяме по A<sub>2</sub>

$$A_2 = T: 3 \text{ примера} \Rightarrow 1+, 2- \Rightarrow \text{Entropy}(S) = -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3} \approx 0,918296$$

$$A_2 = F: 3 \text{ примера} \Rightarrow 1+, 2- \Rightarrow \text{Entropy}(S) = -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3} \approx 0,918296$$

$$\sum = \frac{3}{6} \cdot 0,918296 + \frac{3}{6} \cdot 0,918296 = 0,918296$$

Информационна печалба:

$$\text{Gain}(S, A2) = \text{Entropy}(S) - 0,918296 = 0$$

### Задача 2.

- a) Начертайте класификационното дърво, което ще бъде научено от ID3 алгоритъма след четири обучаващи примера на понятието Харесва, описано в лекция 1 и зададени от следната таблица:

Пример	Небе	Въздух	Влажност	Вятър	Вода	Прогноза	Харесва
1	Сънце	Топъл	Нормална	Силен	Топла	Същото	Да
2	Сънце	Топъл	Висока	Силен	Топла	Същото	Да
3	Дъжд	Студен	Висока	Силен	Топла	Промяна	Не
4	Сънце	Топъл	Висока	Силен	Студена	Промяна	Да

$$\text{Entropy}(S) = -\frac{3}{4} \log_2 \frac{3}{4} - \frac{1}{4} \log_2 \frac{1}{4} \approx 0,811278$$

Информационна печалба за атрибут „Небе“:

<b>Небе</b>	<b>Примери</b>	<b>Класове</b>	<b>Ентропия</b>
Сънце	3	Да, Да, Да	0
Дъжд	1	Не	0

$$Gain(S, \text{Небе}) = 0.811 - \left( \frac{3}{4} \cdot 0 + \frac{1}{4} \cdot 0 \right) = 0.811$$

Информационна печалба за атрибут „Въздух“

<b>Въздух</b>	<b>Примери</b>	<b>Класове</b>	<b>Ентропия</b>
Топъл	3	Да, Да, Да	0
Студен	1	Не	0

$$Gain(S, \text{Въздух}) = 0.811 - \left( \frac{3}{4} \cdot 0 + \frac{1}{4} \cdot 0 \right) = 0.811$$

Информационна печалба за атрибут „Прогноза“

<b>Прогноза</b>	<b>Примери</b>	<b>Класове</b>	<b>Ентропия</b>
Същото	2	Да, Да	0
Промяна	2	Не, Да	1.0

$$Gain(S, \text{Прогноза}) = 0.811 - \left( \frac{2}{4} \cdot 0 + \frac{2}{4} \cdot 1 \right) = 0.811$$

$$Gain(S, \text{Прогноза}) = 0.811 - 0.5 = 0.311$$

Информационна печалба за „Влажност“

<b>Влажност</b>	<b>Примери</b>	<b>Класове</b>	<b>Ентропия</b>
Нормална	1	Да	0
Висока	3	Да, Да, Не	$\left( -\frac{2}{3} \cdot \log_2 \frac{2}{3} - \frac{1}{3} \cdot \log_2 \frac{1}{3} \right) \approx 0.918296 \right)$

$$Gain(S, \text{Влажност}) = 0.811 - \left( \frac{1}{4} \cdot 0 + \frac{3}{4} \cdot 0.918 \right) = 0.123$$

Информационна печалба за „Вятър“

$$Gain(S, \text{Вятър}) = 0.811 - \frac{4}{4} \cdot 0.811 = 0$$

Информационна печалба за „Вода“

### Вода Примери Класове Ентропия

Топла 3 Да, Да, Не 0.918

Студена 1 Да 0

$$Gain(S, \text{Вода}) = 0.811 - \left( \frac{3}{4} \cdot 0.918 + \frac{1}{4} \cdot 0 \right)$$

$$Gain(S, \text{Вода}) = 0.811 - 0.688 = 0.123$$

### Атрибут Информационна печалба

Небе **0.811**

Въздух **0.811**

Прогноза 0.311

Влажност 0.123

Вода 0.123

Вятър 0.000



b) Добавете новия пример:

Пример	Небе	Въздух	Влажност	Вятър	Вода	Прогноза	Харесва
5	Слънце	Топло	Нормална	Слаб	Топла	Същото	не

Постройте ново дърво и укажете стойността на информационната печалба за всеки кандидат атрибут при всяка стъпка от построяването на дървото.

Класове: 3 „Да“ (№1,2,4) и 2 „Не“ (№3,5).

$$Entropy(S) = -0.6\log_2 0.6 - 0.4\log_2 0.4 \approx 0.971.$$

Информационна печалба на корена (всички атрибути)

Стойности на Небе = { Слънце , Дъжд }

### Стойност Положителни Отрицателни Ентропия

$$\text{Слънце } 3 \quad 1 \quad \left( -\frac{3}{4} \cdot \log_2 \frac{3}{4} - \frac{1}{4} \cdot \log_2 \frac{1}{4} \right) \approx 0.811$$

$$\text{Дъжд } 0 \quad 1 \quad \left( -0 \cdot \log_2 0 - \frac{1}{1} \cdot \log_2 \frac{1}{1} \right) = 0$$

$$Gain(S, \text{Небе}) = 0.971 - \left( \frac{4}{5} \cdot 0.811 + \frac{1}{5} \cdot 0 \right) = 0.3222$$

Изчисление за атрибута Влажност:

<b>Влажност</b>	<b>Примери</b>	<b>Класове</b>	<b>Ентропия</b>
Нормална	2	Да, Не	1
Висока	3	Да, Да, Не	$\left(-\frac{2}{3} \cdot \log_2 \frac{2}{3} - \frac{1}{3} \cdot \log_2 \frac{1}{3}\right) \approx 0.918296$

$$Gain(S, \text{Влажност}) = 0.971 - \left( \frac{2}{5} \cdot 1 + \frac{3}{5} \cdot 0.918 \right) = 0.020$$

Информационна печалба за „Вятър“

#### **Стойност Положителни Отрицателни Ентропия**

Силен	3	1	0.811
Слаб	0	1	0

$$Gain(S, \text{Вятър}) = 0.971 - \left( \frac{4}{5} \cdot 0.811 + \frac{1}{5} \cdot 0 \right)$$

$$Gain(S, \text{Вятър}) = 0.971 - 0.649 = 0.3222$$

Информационна печалба за „Вода“

#### **Вода Примери Класове Ентропия**

Топла	3	Да, Да, Не, Не	1
Студена	1	Да	0

$$Gain(S, \text{Вода}) = 0.971 - \left( \frac{4}{5} \cdot 1 + \frac{1}{5} \cdot 0 \right)$$

$$Gain(S, \text{Вода}) = 0.971 - 0.8 = 0.171$$

Информационна печалба за „Прогноза“

<b>Прогноза</b>	<b>Примери</b>	<b>Класове</b>	<b>Ентропия</b>
Същото	3	Да, Да, Не	$\left(-\frac{2}{3} \cdot \log_2 \frac{2}{3} - \frac{1}{3} \cdot \log_2 \frac{1}{3}\right) \approx 0.918296$
Промяна	2	Да, Не	1

$$Gain(S, \text{Прогноза}) = 0.971 - \left( \frac{3}{5} \cdot 0.918 + \frac{2}{5} \cdot 1 \right)$$

$$Gain(S, \text{Прогноаз}) = 0.971 - 0.951 = 0.020$$

Има равенство (Небе, Въздух, Вятър) с  $IG \approx 0.323$ . Ще изберем Небе.

Второ ниво: подмножеството *Небе=Слънце*

Тук имаме 4 примера: (3 Да, 1 Не), Entropy=0.811.

- Въздух: всички са Топъл  $\rightarrow$  IG=0
- Влажност: Нормална $\rightarrow$ (1,1) Ентропия=1; Висока $\rightarrow$ (2,0) Ентропия=0

$$Gain = 0.811 - \left( \frac{2}{4} \cdot 1 + \frac{2}{4} \cdot 0 \right) = \mathbf{0.311}$$

- Вятър: Силен $\rightarrow$ (3,0) Ентропия=0; Слаб $\rightarrow$ (0,1) Ентропия=0

$$Gain = \mathbf{0.811}$$

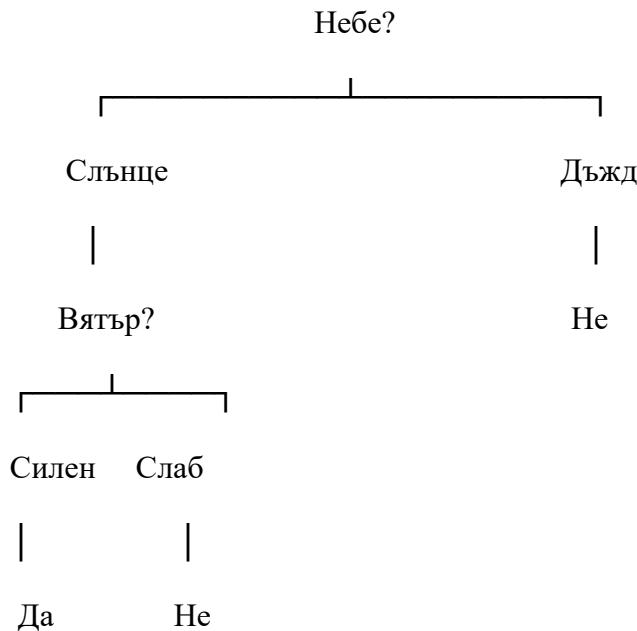
- Вода: Топла $\rightarrow$ (2,1) Ентропия=0.918; Студена $\rightarrow$ (1,0) Ентропия=0

$$Gain = 0.811 - \left( \frac{3}{4} \cdot 0.918 + \frac{1}{4} \cdot 0 \right) = \mathbf{0.123}$$

- Прогноза: Същото $\rightarrow$ (2,1) Ентропия=0.918; Промяна $\rightarrow$ (1,0) Ентропия=0

$$Gain = 0.811 - \left( \frac{3}{4} \cdot 0.918 + \frac{1}{4} \cdot 0 \right) = \mathbf{0.123}$$

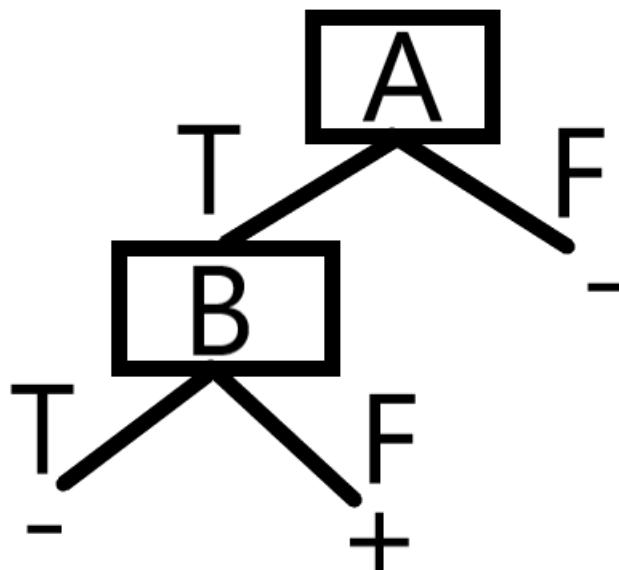
Избор за възел под „Слънце“: *Вятър*.



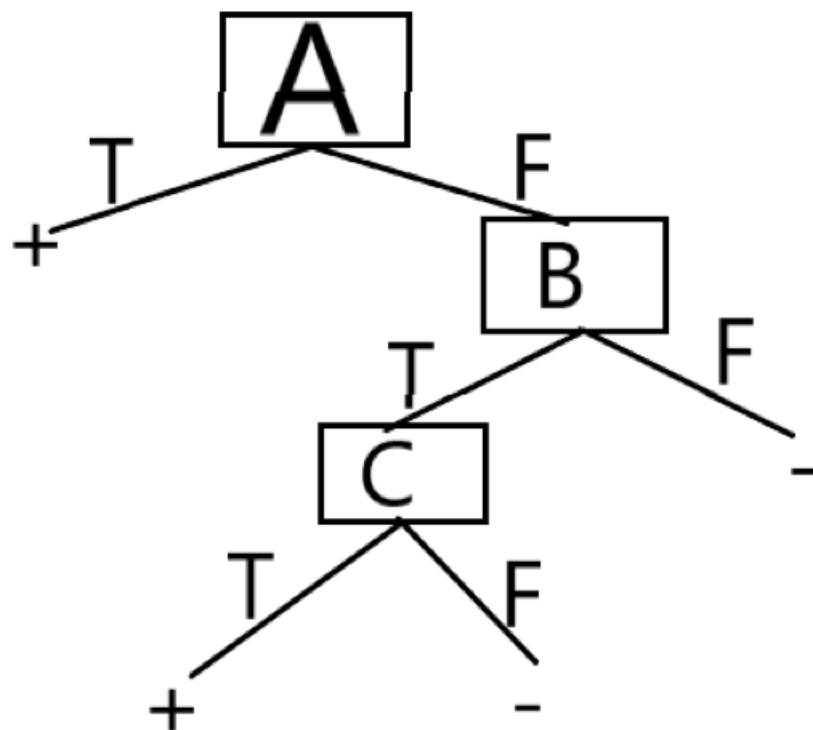
**Задача 3.**

Нарисувайте класификационни дървета, представящи следните Булеви функции:

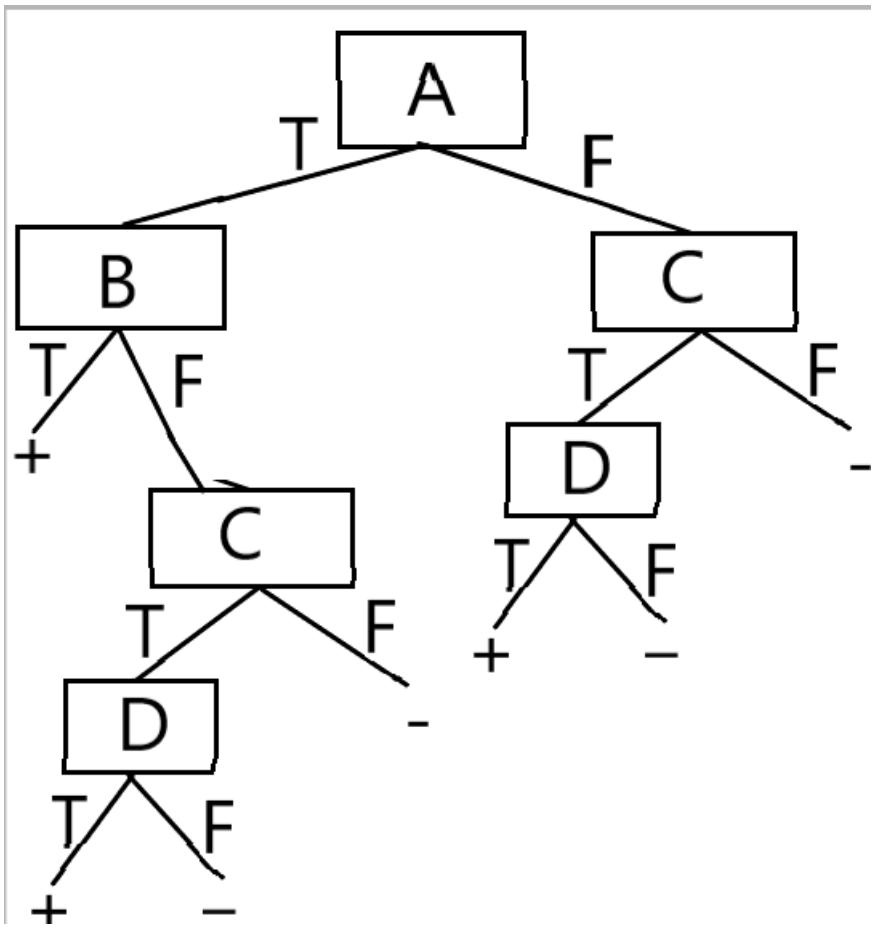
a)  $A \wedge \neg B$



b)  $A \vee (B \wedge C)$



c)  $(A \wedge B) \vee (C \wedge D)$



#### Задача 4.

Представете, че има 2 класификационни дървета D1 и D2, описващи някоя булева функция. Дървото D2 е направено чрез нарастване на дърво D1 (т.е. един от терминалните възли на D е заменен с поддърво).

Дали следващото утвърждение е вярно:

Дърво D1 е по-общо-или-равно-на дърво D2?

Ако „да“ – докаже го. Ако „не“ – дайте контра-пример.

Имаме две дървета:

- $D_1$ : никакво класификационно дърво
- $D_2$ : получено от  $D_1$ , като **един от листовете на  $D_1$**  е заменен с поддърво

е променяме останалите части на  $D_1$ , а само един лист — разклоняваме го на нов тест.

#### Пример 1 – Разрастваме положителен лист (+)

A

/ \

T F

+ -

След разрастване на положителния лист (+):

A

/ \

T F

|

B

/ \

T F

+ -

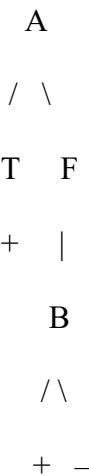
Получаваме:

- $A=T$  и  $B=T \rightarrow +$
- $A=T$  и  $B=F \rightarrow -$
- $A=F \rightarrow -$

Всички примери, които преди са положителни в  $D_2$ , вече са били положителни и в  $D_1$  (защото в  $D_1$  при  $A=T$  имаме „+“).

Следователно  $D_1$  е по-общо или равно на  $D_2$ .

### Пример 2 – Разрастваме отрицателен лист (-)



Новото поддърво добавя положителни изходи в клон, който преди винаги беше „–“. Тоест се появяват примери, при които  $D_2(x)=1$ , а  $D_1(x)=0$ .

Следователно импликацията  $D_2(x)=1 \Rightarrow D_1(x)=1$  се нарушава.

Ако  $D_2$  е направено от  $D_1$  чрез нарастване на лист, не можем автоматично да кажем, че  $D_1$  е по-общо. Това зависи от знака на листа, който е разраснат.

- Ако е бил  $+$ , тогава  $D_1$  е по-общо.
- Ако е бил  $-$ , тогава  $D_2$  става по-общо.