

Машинное обучение  
Практические задания №4

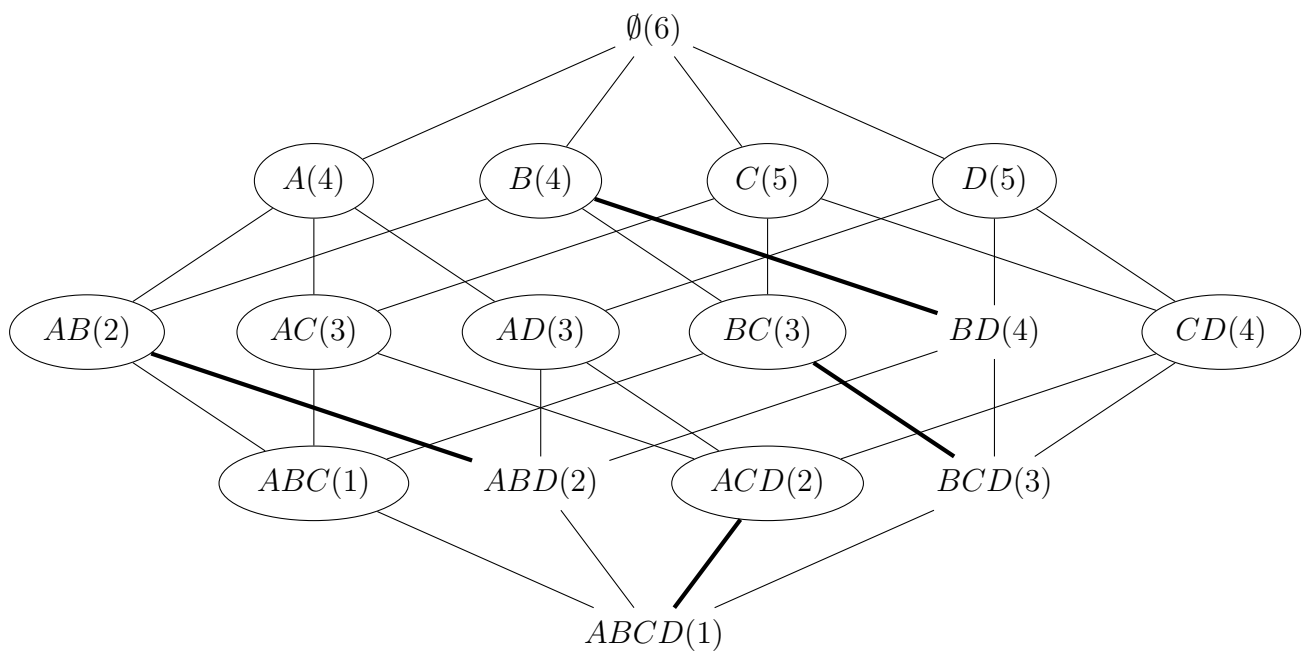
Задача 1. Дан набор данных

<i>Tid</i>	<i>Itemset</i>
1	<i>ACD</i>
2	<i>BCD</i>
3	<i>AC</i>
4	<i>ABD</i>
5	<i>ABCD</i>
6	<i>BCD</i>

Найдите все минимальные генераторы для минимального уровня поддержки = 1.

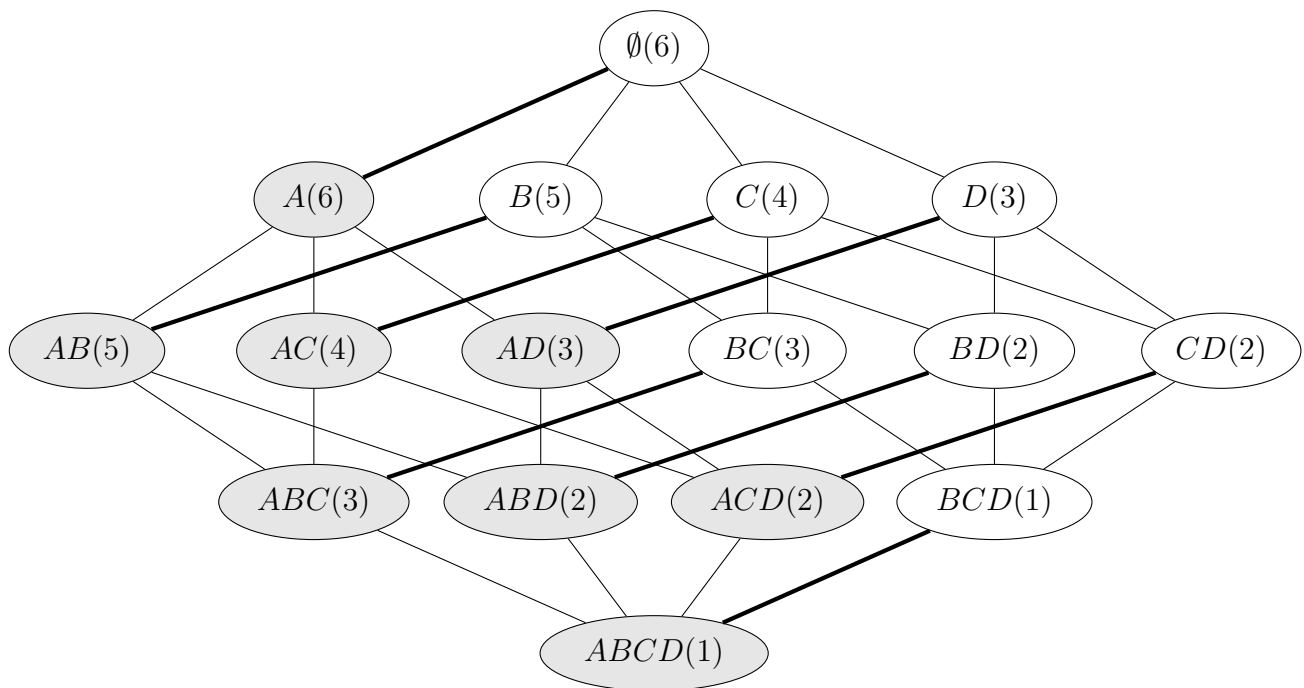
Множество является минимальным генератором, когда оно не имеет подмножеств с тем же уровнем поддержки.

Решение. Построим решетку наборов и их частот.



Наборы, которые являются минимальными генераторами обведены. Жирным выделены ребра, которые указывают на подмножества с такой же поддержкой. Множество минимальных генераторов:  $\{A, B, C, D, AB, AC, AD, BC, CD, ABC, ACD\}$ .  $\square$

**Задача 2.** Дана решетка наборов и их частоты.



Выполните следующие задания:

- А. Выпишите список всех замкнутых наборов (*closed itemsets*). Множество замкнутое, когда нет надмножеств с тем же уровнем поддержки.
- В. Является ли набор  $BCD$  выводимым? Является ли набор  $ABCD$  выводимым? Какие границы их поддержки?

Решение.

- А. На графе выделены ребра, которые ведут к надмножествам с такой же поддержкой. Множество замкнутых наборов:  $\{A, AB, AC, AD, ABC, ABD, ACD, ABCD\}$ .

В.

$sup(BCD)$	$\geq 0$	$Y = BCD$
	$\leq sup(BC) = 3$	$Y = BC$
	$\leq sup(BD) = 2$	$Y = BD$
	$\leq sup(CD) = 2$	$Y = CD$
	$\geq sup(BC) + sup(BD) - sup(B) = 0$	$Y = B$
	$\geq sup(BC) + sup(CD) - sup(C) = 0$	$Y = C$
	$\geq sup(BD) + sup(CD) - sup(D) = 0$	$Y = D$
	$\leq sup(BC) + sup(BD) + sup(CD) - sup(B) - sup(C) - sup(D) + sup(\emptyset) = 1$	$Y = \emptyset$

$$\text{LB}(BCD) = \{0, 1\}, \text{UB}(BCD) = \{1, 2, 3\}$$

$$\max(\text{LB}(BCD)) = \min(\text{UB}(BCD)) = 1, \Rightarrow \text{набор выводим.}$$

$\text{sup}(ABCD)$	$\geq 0$	$Y = ABCD$
	$\leq \text{sup}(ABC) = 3$	$Y = ABC$
	$\leq \text{sup}(ABD) = 2$	$Y = ABD$
	$\leq \text{sup}(ACD) = 2$	$Y = ACD$
	$\leq \text{sup}(BCD) = 1$	$Y = BCD$
	$\geq \text{sup}(ABC) + \text{sup}(ABD) - \text{sup}(AB) = 0$	$Y = AB$
	$\geq \text{sup}(ABC) + \text{sup}(ACD) - \text{sup}(AC) = 1$	$Y = AC$
	$\geq \text{sup}(ABD) + \text{sup}(ACD) - \text{sup}(AD) = 1$	$Y = AD$
	$\geq \text{sup}(ABC) + \text{sup}(BCD) - \text{sup}(BC) = 1$	$Y = BC$
	$\geq \text{sup}(ABD) + \text{sup}(BCD) - \text{sup}(BD) = 1$	$Y = BD$
	$\geq \text{sup}(ACD) + \text{sup}(BCD) - \text{sup}(CD) = 1$	$Y = CD$
	$\leq \text{sup}(ABC) + \text{sup}(ABD) + \text{sup}(ACD) -$ $-\text{sup}(AB) - \text{sup}(AC) - \text{sup}(AD) + \text{sup}(A) = 1$	$Y = A$
	$\leq \text{sup}(ABC) + \text{sup}(ABD) + \text{sup}(BCD) -$ $-\text{sup}(AB) - \text{sup}(BC) - \text{sup}(BD) + \text{sup}(B) = 1$	$Y = B$
	$\leq \text{sup}(ABC) + \text{sup}(ACD) + \text{sup}(BCD) -$ $-\text{sup}(AC) - \text{sup}(BC) - \text{sup}(CD) + \text{sup}(C) = 1$	$Y = C$
	$\leq \text{sup}(ABD) + \text{sup}(ACD) + \text{sup}(BCD) -$ $-\text{sup}(AD) - \text{sup}(BD) - \text{sup}(CD) + \text{sup}(D) = 1$	$Y = D$
	$\geq \text{sup}(ABC) + \text{sup}(ABD) + \text{sup}(ACD) + \text{sup}(BCD) -$ $-\text{sup}(AB) - \text{sup}(AC) - \text{sup}(AD) - \text{sup}(BC) -$ $-\text{sup}(BD) - \text{sup}(CD) + \text{sup}(A) + \text{sup}(B) + \text{sup}(C) +$ $+\text{sup}(D) - \text{sup}(\emptyset) = 1$	$Y = \emptyset$

$$\text{LB}(ABCD) = \{0, 1\}, \text{UB}(ABCD) = \{1, 2, 3\}$$

$$\max(\text{LB}(ABCD)) = \min(\text{UB}(ABCD)) = 1, \Rightarrow \text{набор выводим.}$$

□

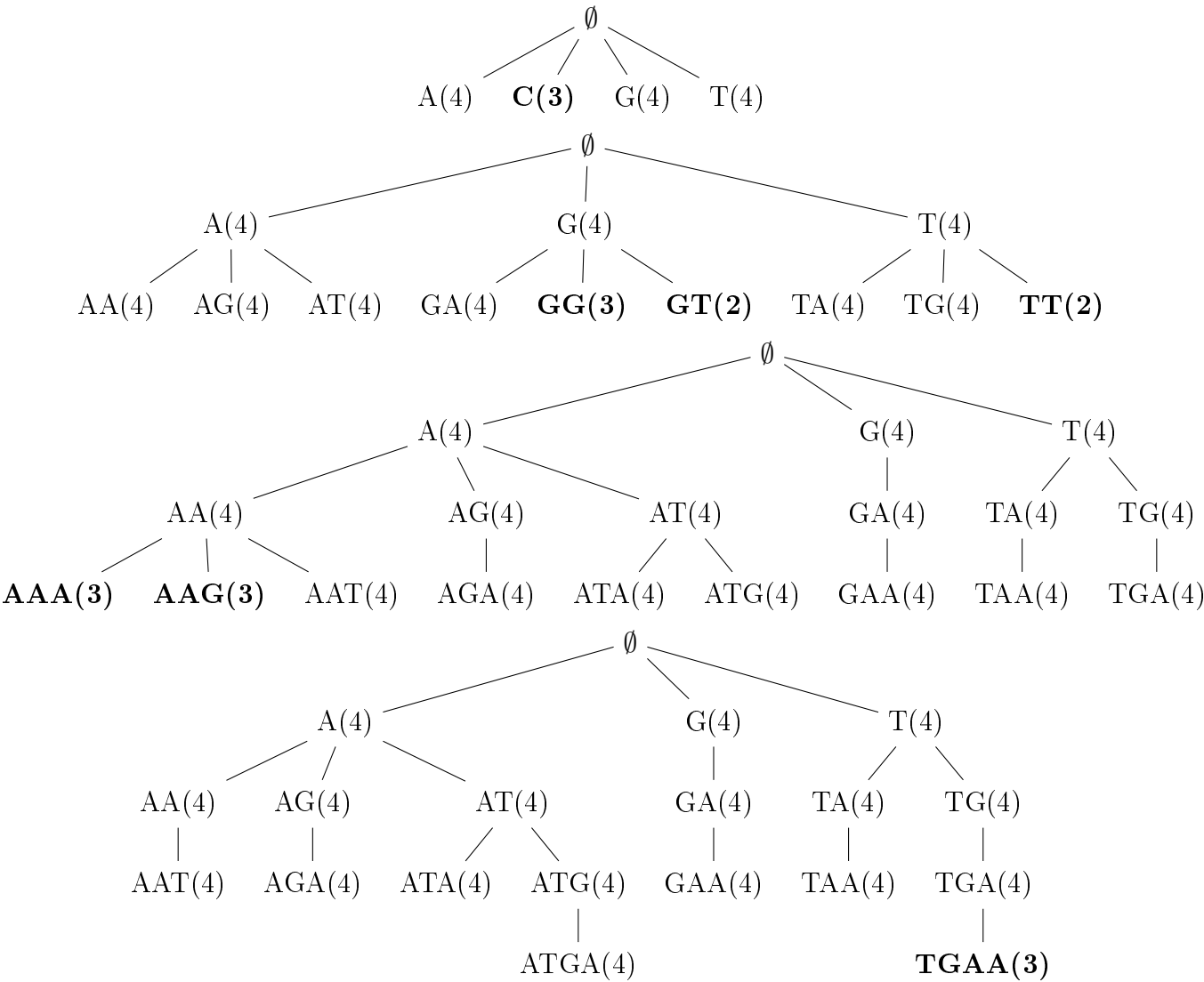
**Задача 3.** Для алфавита  $\{A, C, G, T\}$  посчитайте, сколько всего может быть разных последовательностей длины  $k$ .  
 Даны последовательности

<i>Id</i>	<i>Sequence</i>
$s_1$	AATACAAGAAC
$s_2$	GTATGGTGAT
$s_3$	AACATGGCCAA
$s_4$	AAGCGTGGTCAA

Найдите все подпоследовательности с минимальным уровнем поддержки  $= 4$ .

Решение. Возможных последовательностей длины  $k$  может быть  $4^k$ .

Для поиска всех подпоследовательностей построим деревья, где каждое следующее дерево является продолжением предыдущего.



□