

Алгоритм кластеризації k-means (1)

1

Задано набір з 8 точок у двовимірному просторі, який треба розбити на два кластери:

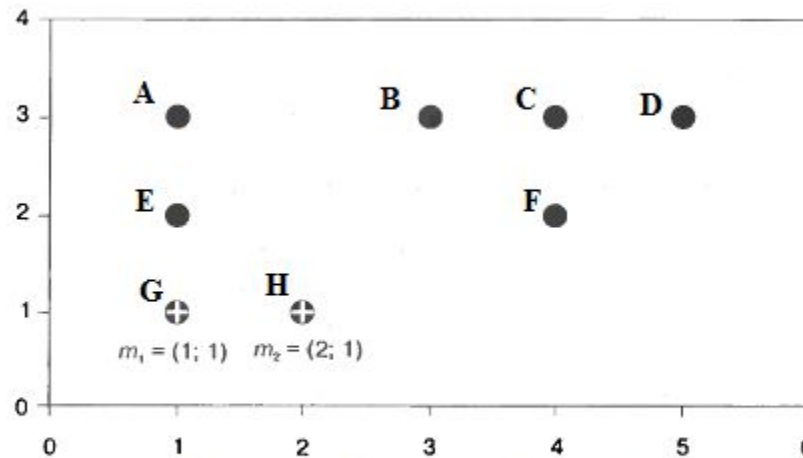
A	B	C	D	T	F	G	H
(1;3)	(3;3)	(4;3)	(5;3)	(1;2)	(4;2)	(1;1)	(2;1)

Крок 1. Визначимо кількість кластерів, на яку треба розбити початкову множину: $k=2$.

Крок 2. Випадковим чином визначимо дві точки $m_1=G$ і $m_2=H$, як центри кластерів.

Крок 3, прохід 1. Для кожної точки визначимо найближчий до неї центр кластеру у евклідовій метриці, тим самим визначаючи, до якого кластеру вона відноситься.

$$d_2(X,Y) = \sqrt{\sum_i (x_i - y_i)^2}.$$



Точка	A	B	C	D	E	F	G	H
Відстань до m_1	2,00	2,83	3,61	4,47	1,00	3,16	0,00	1,00
Відстань до m_2	2,24	2,24	2,83	3,61	1,41	2,24	1,00	0,00
Кластер	1	2	2	2	1	2	1	2

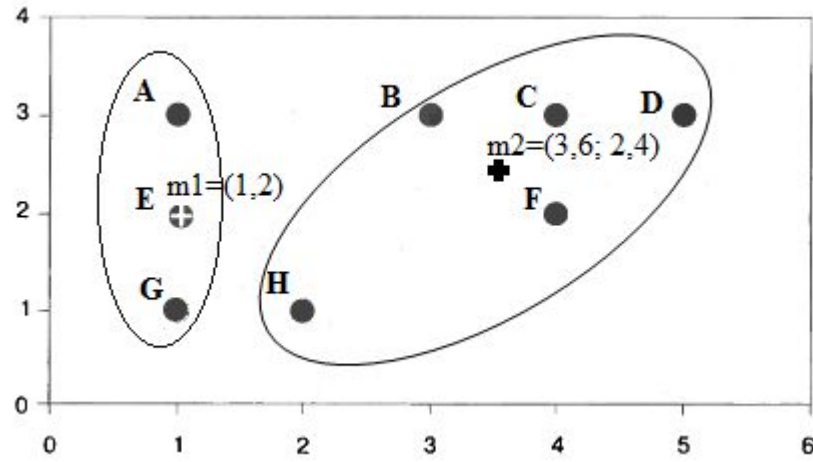
Визначивши належність точок кластерам, обчислюємо суму квадратів помилок:

$$E = s = 1 \sum_{i=1}^k \sum_{p \in C} (p - m)^2 = 2^2 + 2,24^2 + 2,83^2 + 3,61^2 + 1^2 + 2,24^2 + 0^2 + 0^2 = 36;$$

Алгоритм кластеризації k-means (2)

2

Крок 4, прохід 1. Обчислюємо центроїди, до яких переміщуються центр кластерів:
 $\zeta_1 = [(1+1+1/3); (3+2+1/3)] = (1,2)$; $\zeta_2 = [(3+4+5+4+2/5); (3+3+3+2+1/5)] = (3,6; 2,4)$.



Крок 3, прохід 2. Для кожної точки знов визначається найближчий до неї центр нових кластерів і відповідна належність її до цього кластеру:

Точка	A	B	C	D	E	F	G	H
Відстань до m1	1,00	2,24	3,16	4,12	0,00	3,00	1,00	1,41
Відстань до m2	2,67	0,85	0,72	1,52	2,63	0,57	2,95	2,13
Кластер	1	2	2	2	1	2	1	1

Бачимо, що відносно велика зміна значення m2 призвела до того, що точка H стала ближче до центру m1 ставши членом кластеру 1. Нова сума квадратів помилок склала:

$$E = s = 1 \sum_{i=1}^k \sum_{p \in C} (p - m)^2 = 1^2 + 0,85^2 + 0,72^2 + 1,52^2 + 0^2 + 0,57^2 + 1^2 + 1,41^2 = 7,86$$

Помилка зменшилась, що означає краще групування об'єктів відносно центрів кластерів.

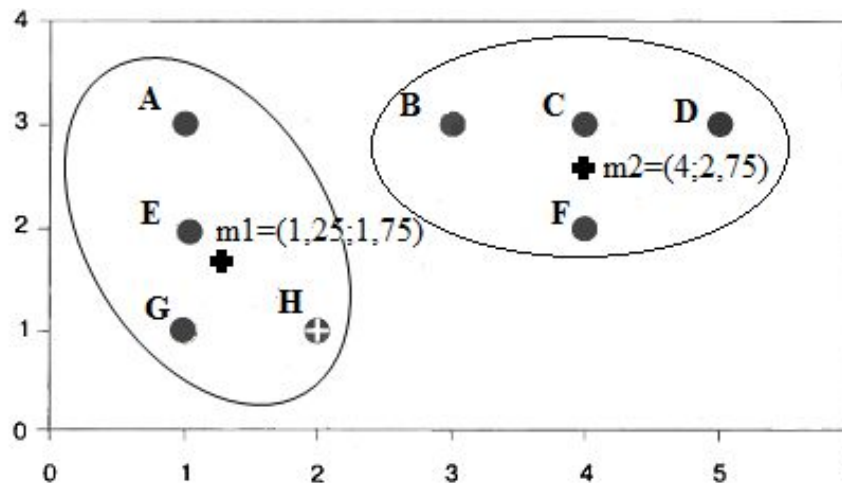
Алгоритм кластеризації k-means (3)

3

Крок 4, прохід 2. Обчислюємо нові центроїди для кожного кластеру:

$\zeta_1 = [(1+1+1+2/4);(3+2+1+1/4)] = (1,25;1,75)$; $\zeta_2 = [(3+4+5+4/4);(3+3+3+2+4)] = (4;2,75)$.

У порівнянні з минулим проходом центри кластерів мало змінилася.



Крок 3, прохід 3. Визначаємо відстані точок від олижчого з центрів нових кластерів:

Точка	A	B	C	D	E	F	G	H
Відстань до m1	1,27	2,15	3,92	3,95	0,35	2,76	0,79	1,06
Відстань до m2	3,01	1,03	0,25	1,03	3,09	0,75	3,47	2,66
Кластер	1	2	2	2	1	2	1	1

нова сума квадратів помилок склала:

$$E = s = 1 \sum_{i=1}^k \sum_{p \in C} (p - m)^2 = 1,27^2 + 1,03^2 + 0,25^2 + 1,03^2 + 0,35^2 + 0,75^2 + 0,79^2 + 1,06^2 = 6,23$$

Сума квадратів помилок мала змінилась відносно попереднього проходу.

Крок 4, прохід 3. Обчислюємо нові центроїди кластерів. Оскільки жодний об'єкт не змінив свого членства у кластерах і положення центроїдів практично не змінилося, алгоритм завершує свою роботу.

Наївний Байєсовський класифікатор (1)

4

Для заданого набору даних, з використанням наївного байєсовського класифікатора визначте, який статус ймовірно має особа (а) з відділу маркетингу у віці від 31 до 35 років, з зарплатою від 46 до 50 тисяч на місяць; (б) з відділу продажів у віці від 31 до 35 років, з зарплатою від 66 до 70 тисяч на місяць

Відділ	продаж	продаж	Продаж	систем	систем	систем	систем	маркет	маркет	секрет	секрет
Вік	31-35	26-30	31-35	21-25	31-35	26-30	41-45	36-40	31-35	46-50	26-30
Зарплата	46-50	26-30	31-35	46-50	66-70	46-50	66-70	46-50	41-45	36-40	26-30
Статус	старш	молод	Молод	Молод	старш	молод	старш	старш	молод	старш	молод

$$P(A \& B) = P(A)P(B|A) = P(B)P(A|B); \quad P(B|A) = P(B)P(A|B)/P(A);$$

Відповідь:

$$A) P(\text{старш}|\text{маркет}; 31-35; 46-50) = P(\text{старш}) P(\text{маркет}|\text{старш})$$

$$P(31-35|\text{старш}) P(46-50|\text{старш}) = 5/11 \times 1/5 \times 2/5 \times 2/5 = 0,0145$$

$$P(\text{молод}|\text{маркет}; 31-35; 46-50) = P(\text{молод}) P(\text{маркет}|\text{молод}) P(31-35|\text{молод})$$

$$P(46-50|\text{молод}) = 6/11 \times 1/6 \times 2/6 \times 2/6 = 0,0101$$

$$P(\text{старш}|\text{маркет}; 31-35; 46-50) = 0,0145 / (0,0145 + 0,0101) = 0,0145 / 0,0246 = 0,59$$

$$P(\text{молод}|\text{маркет}; 31-35; 46-50) = 0,0101 / (0,0145 + 0,0101) = 0,0101 / 0,0246 = 0,41$$

$$B) P(\text{старш}|\text{продаж}; 31-35; 66-70) = P(\text{старш}) P(\text{продаж}|\text{старш}) P(31-35|\text{старш}) P(66-70|\text{старш}) = 5/11 \times 1/5 \times 2/5 \times 2/5 = 0,006$$

$$P(\text{молод}|\text{продаж}; 31-35; 76-70) = P(\text{молод}) P(\text{продаж}|\text{молод}) P(31-35|\text{молод}) P(66-70|\text{молод}) = 6/11 \times 2/6 \times 2/6 \times 0/6 = 0$$

$$P(\text{старш}|\text{продаж}; 31-35; 66-70) = 1$$

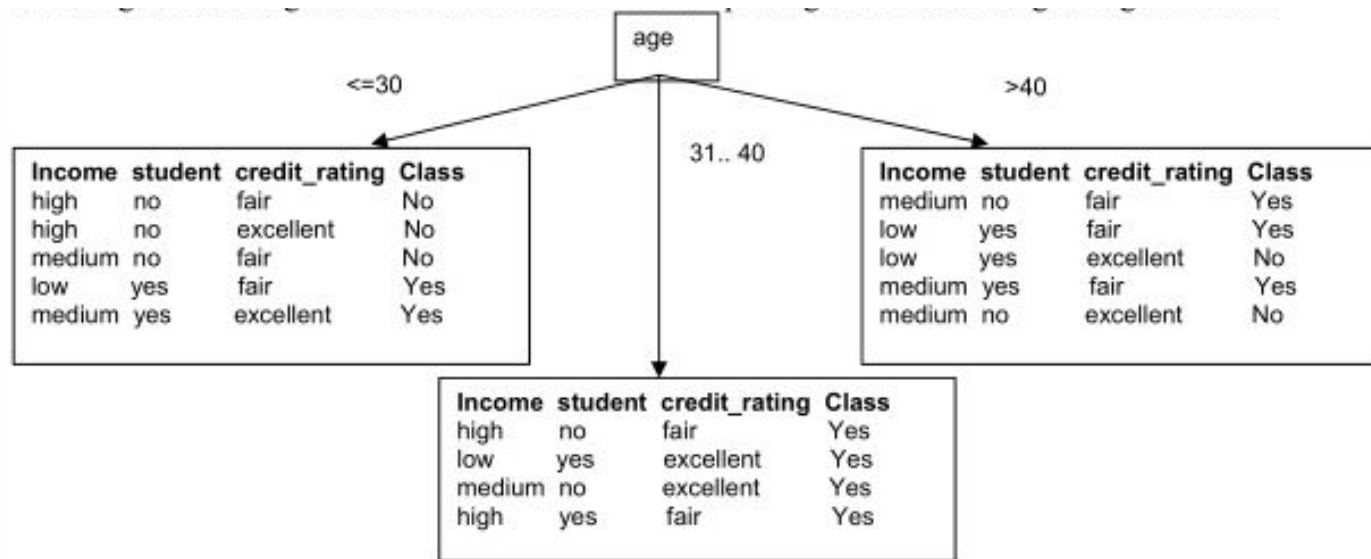
$$P(\text{молод}|\text{продаж}; 31-35; 76-70) = 0$$

ДЕРЕВА РІШЕНЬ (1)

5

На основі навчальної вибірки побудуйте дерево рішень для визначення бажання різних категорій споживачів щодо купівлі комп'ютера

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Вік	<=30	<=30	31-40	>40	>40	>40	31-40	<=30	<=30	>40	<=30	31-40	31-40	>40
Дохід	висок	висок	висок	серед	низьк	низьк	низьк	серед	Низьк	Серед	серед	серед	висок	серед
Студ	ні	ні	ні	ні	так	Так	так	Ні	так	так	так	ні	так	ні
кредит	гарна	відм	гарна	гарна	гарна	відм	відм	гарна	гарна	гарна	відм	відм	гарна	Відм
купить	ні	ні	так	так	так	Ні	так	Ні	так	так	так	так	так	ні



$$I(S_{\text{ТАК}}, S_{\text{НІ}}) = I(9,5) = -9/14 \log_2(9/14) - 5/14 \log_2(5/14) = 0.94$$

Вік: 3 значення: **<=30** (2 так,3 ні), **31..40** (4 так,0 ні), **>40** (3 так,2 ні)

ДЕРЕВА РІШЕНЬ (2)

6

На основі навчальної вибірки побудуйте дерево рішень для визначення бажання різних категорій споживачів щодо купівлі комп'ютера

№	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Вік	<=30	<=30	31-40	>40	>40	>40	31-40	<=30	<=30	>40	<=30	31-40	31-40	>40
Дохід	висок	висок	висок	серед	низьк	низьк	низьк	серед	Низьк	Серед	серед	серед	висок	серед
Студ	ні	ні	ні	ні	так	Так	так	Ні	так	так	так	ні	так	ні
кредит	гарна	відм	гарна	гарна	гарна	відм	відм	гарна	гарна	гарна	відм	відм	гарна	Відм
купить	ні	ні	так	так	так	Ні	так	Ні	так	так	так	так	так	ні

$$I(S_{\text{ТАК}}, S_{\text{НІ}}) = I(9,5) = -9/14 \log(9/14) - 5/14 \log(5/14) = 0.94$$

Вік: 3 значення: <=30 (2 так, 3 ні), 31..40 (4 так, 0 ні), >40 (3 так, 2 ні)

$$\text{Entropy(вік)} = 5/14 (-2/5 \log(2/5) - 3/5 \log(3/5)) + 4/14 (0) + 5/14 (-3/5 \log(3/5) - 2/5 \log(2/5)) = 5/14(0.9709) + 0 + 5/14(0.9709) = 0.6935$$

$$\text{Gain(age)} = 0.94 - 0.6935 = 0.2465$$

Дохід 3 значення: високий (2 так, 2 ні), середній (4 так, 2 ні), низький (3 так, 1 ні)

$$\text{Entropy(дохід)} = 4/14 (-2/4 \log(2/4) - 2/4 \log(2/4)) + 6/14 (-4/6 \log(4/6) - 2/6 \log(2/6))$$

$$+ 4/14 (-3/4 \log(3/4) - 1/4 \log(1/4)) = 4/14 (1) + 6/14 (0.918) + 4/14 (0.811) =$$

$$0.285714 + 0.393428 + 0.231714 = 0.9108 \text{ Gain(дохід)} = 0.94 - 0.9108 = 0.0292$$

Студент: 2 значення: так (6 так, 1 ні), ні (3 так, 4 ні)

$$\text{Entropy(студент)} = 7/14 (-6/7 \log(6/7)) + 7/14 (-3/7 \log(3/7) - 4/7 \log(4/7)) = 7/14(0.5916) + 7/14(0.9852) = 0.2958 + 0.4926 = 0.7884$$

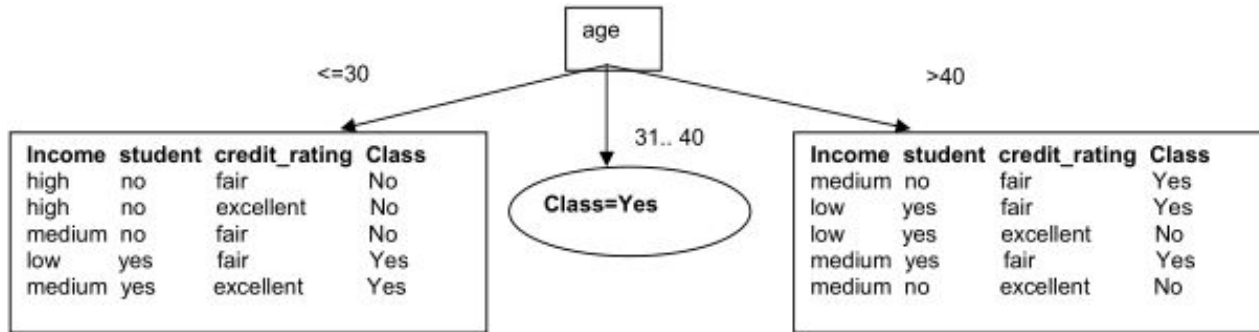
$$\text{Gain(студент)} = 0.94 - 0.7884 = 0.1516$$

Кредит: 2 значення: гарна (6 так, 2 ні), відмінна (3 так, 2 ні)

$$\text{Entropy(кредит)} = 8/14 (-6/8 \log(6/8) - 2/8 \log(2/8)) + 6/14 (-3/6 \log(3/6) - 3/6 \log(3/6)) = 8/14(0.8112) + 6/14(1) = 0.4635 + 0.4285 = 0.8920$$

$$\text{Gain(кредит)} = 0.94 - 0.8920 = 0.0480$$

ДЕРЕВА РІШЕНЬ (3)



Ентропія блоку: $I(\text{СТАК}, \text{SHI}) = I(2, 3) = -2/5 \log(2/5) - 3/5 \log(3/5) = 0.97$
 Дохід: 3 значення: **високий** (0 так, 2 ні), **середній** (1 так, 1 ні), **низький** (1 так, 0 ні)

Entropy(дохід) = $2/5(0) + 2/5(-1/2\log(1/2) - 1/2\log(1/2)) + 1/5(0) = 2/5(1) = 0.4$

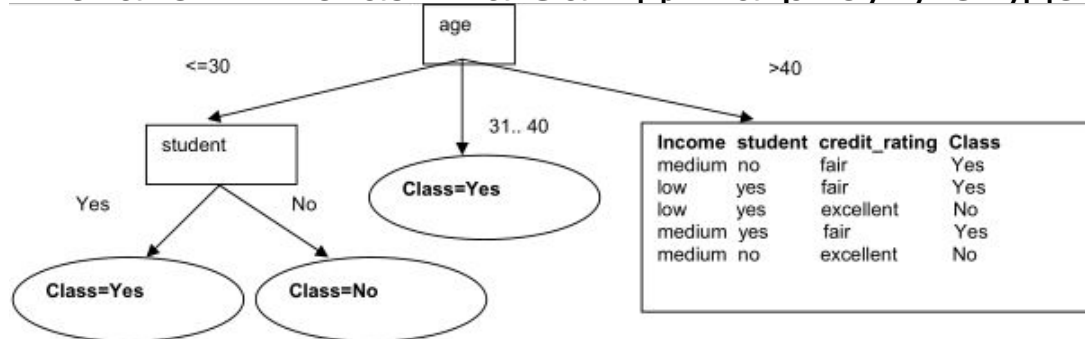
Gain(дохід) = $0.97 - 0.4 = 0.57$

Студент: 2 значення: **так** (2 так, 0 ні), **ні** (0 так, 3 ні)

Entropy(студент) = $2/5(0) + 3/5(0) = 0$

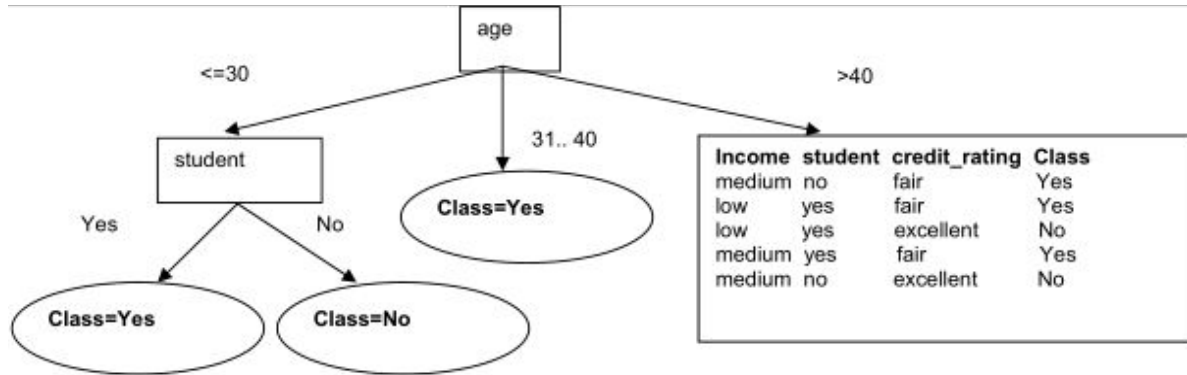
Gain(student) = $0.97 - 0 = 0.97$

Можна робити розбиття по атрибуту студент без перевірки інших атрибутів, оскільки значення показника Gain для атрибуту Студент є максимальним



ДЕРЕВА РІШЕНЬ (4)

8



Ентропія блоку: $I(\text{СТАК}, \text{SHI}) = I(3,2) = -3/5 \log(3/5) - 2/5 \log(2/5) = 0.97$

Дохід: 2 значення: **середній** (2 так, 1 ні), **низький** (1 так, 1 ні)

Entropy(дохід) = $3/5(-2/3 \log(2/3) - 1/3 \log(1/3)) + 2/5(-1/2 \log(1/2) - 1/2 \log(1/2)) = 3/5(0.9182) + 2/5(1) = 0.55 + 0.4 = 0.95$ **Gain(income) = 0.97 - 0.95 = 0.02**

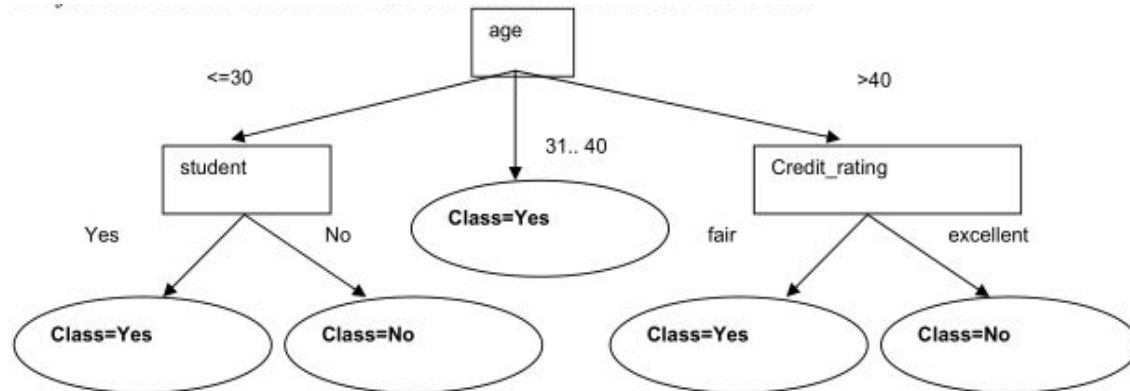
Студент: 2 значення: **так** (2 так, 1 ні), **ні** (1 так, 1 ні)

Entropy(студент) = $3/5(-2/3 \log(2/3) - 1/3 \log(1/3)) + 2/5(-1/2 \log(1/2) - 1/2 \log(1/2)) = 0.95$ **Gain(student) = 0.97 - 0.95 = 0.02**

Кредит: 2 значення: **гарна** (3 так, 0 ні), **відмінна** (0 так, 2 ні)

Entropy(кредит) = 0 **Gain(кредит) = 0.97 - 0 = 0.97**

Здійснюємо розбиття по атрибуту КРЕДИТ, яке дасть два чисті класи:

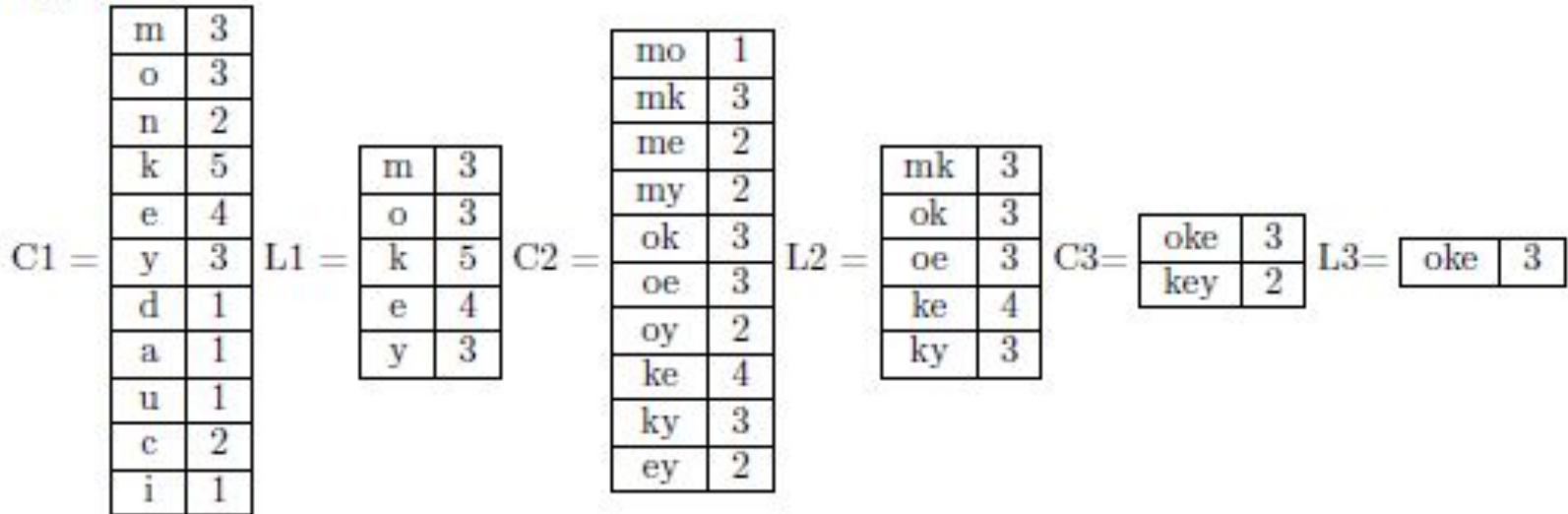


АСОЦІАТИВНІ ПРАВИЛА (1)

9

T1{M,O,N,K,E,Y}; T2{D,O,N,K,E,Y}; T3{M,A,K,E}; T4{{M,U,C,K,Y};
T5{C,O,O,K,I,E}; підтримка – 60%; довіра – 80%.

Apriori:



ПРАВИЛА: $A \rightarrow B: P(B|A) = |B \cap A| / |A|$ o,k \rightarrow e [0,6;1]; o,e \rightarrow k [0,6;1];
k,e \rightarrow o [0,6;0,75]

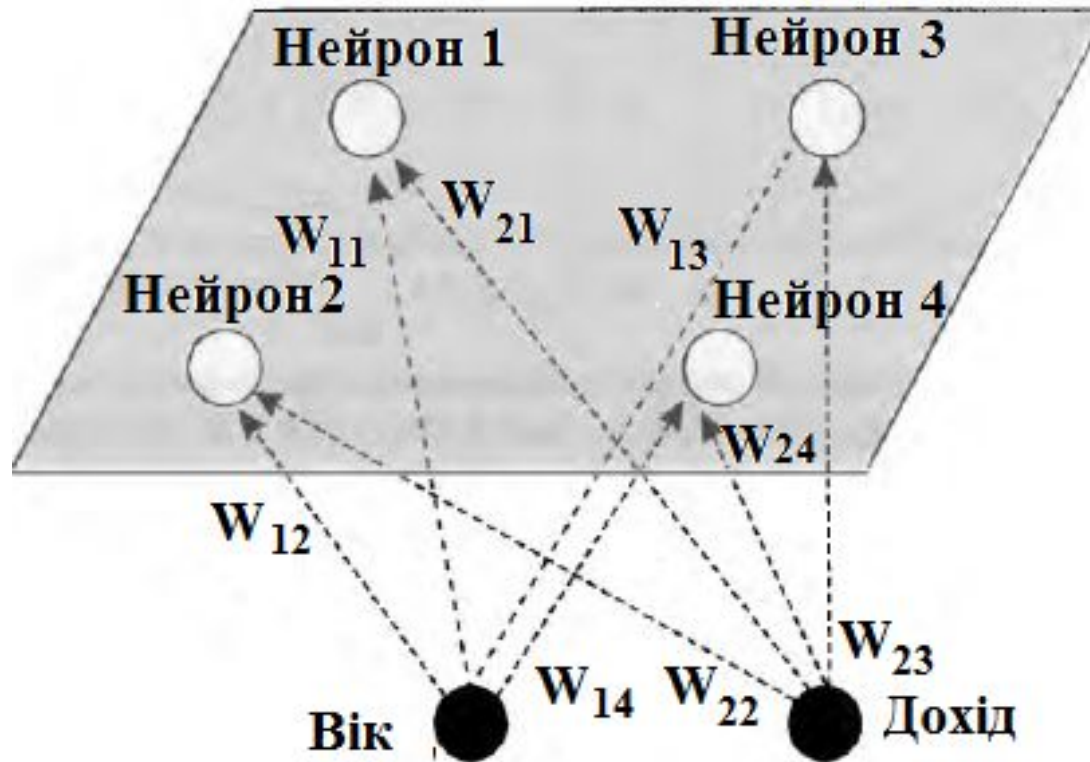
m \rightarrow k [0,6;1]; k \rightarrow m [0,6;0,6] o \rightarrow k [0,6;1] k \rightarrow o [0,6;0,6] o \rightarrow e [0,6;1];
e \rightarrow o [0,6;0,75];

y \rightarrow k [0,6;1]

Відповідь: o \rightarrow k,e [0,6;1]; o,k \rightarrow e [0,6;1]; o,e \rightarrow k [0,6;1]; m \rightarrow k [0,6;1]; o \rightarrow k
[0,6;1];

o \rightarrow e [0,6;1]; y \rightarrow k [0,6;1]

Розглянемо приклад роботи мережі Кохонена, що містить 2 x 2 нейрона у вихідному шарі, а множина даних представлена атрибутами Вік і Дохід з попередньо нормалізованими даними. У зв'язку з малим розміром мережі встановимо радіус навчання $R=0$, тобто можливість підстроювати ваги буде надаватися лише нейрону-переможцю. Коефіцієнт швидкості навчання встановимо $\eta=0,5$.



Випадковим чином виберемо початкові значення ваг нейронів:

w_{11}	w_{21}	w_{12}	w_{22}	w_{13}	w_{23}	w_{14}	w_{24}
0,9	0,8	0,9	0,2	0,1	0,8	0,1	0,2

Сформуємо набір записів вхідної вибірки:

№	x_{1j}	x_{1j}	Опис
1	$x_{11}=0,8$	$x_{12}=0,8$	Літня людина з високим доходом
2	$x_{21}=0,8$	$x_{22}=0,1$	Літня людина з низьким доходом
3	$x_{31}=0,2$	$x_{32}=0,8$	Молода людина з високим доходом
4	$x_{41}=0,1$	$x_{42}=0,2$	Молода людина з низьким доходом

Конкуренція. Обчислимо евклідову відстань між вхідним вектором X_1 і векторами ваг усіх чотирьох нейронів вихідного шару.

$$\begin{aligned} \text{Нейрон 1: } D(\mathbf{W}_1, \mathbf{X}_1) &= \sqrt{(w_{11} - x_{11})^2 + (w_{21} - x_{12})^2} = \sqrt{(0,9 - 0,8)^2 + (0,8 - 0,8)^2} = 0,1. \\ \text{Нейрон 2: } D(\mathbf{W}_2, \mathbf{X}_1) &= \sqrt{(w_{12} - x_{11})^2 + (w_{22} - x_{12})^2} = \sqrt{(0,9 - 0,8)^2 + (0,2 - 0,8)^2} = 0,61. \\ \text{Нейрон 3: } D(\mathbf{W}_3, \mathbf{X}_1) &= \sqrt{(w_{13} - x_{11})^2 + (w_{23} - x_{12})^2} = \sqrt{(0,1 - 0,8)^2 + (0,8 - 0,8)^2} = 0,7. \\ \text{Нейрон 4: } D(\mathbf{W}_4, \mathbf{X}_1) &= \sqrt{(w_{14} - x_{11})^2 + (w_{24} - x_{12})^2} = \sqrt{(0,1 - 0,8)^2 + (0,2 - 0,8)^2} = 0,92. \end{aligned}$$

Переміг нейрон 1, який формує кластер для захоплення літніх людей з високим доходом

Об'єднання. Оскільки радіус навчання дорівнює нулю, тільки нейрон-переможець буде нагороджений можливістю підстроювання свого вектора ваг.

Підстроювання. Для першого нейрона отримуємо формулу:

$$\begin{aligned} w_{i1}^{\text{нове}} &= w_{i1}^{\text{поточне}} + \eta (x_{1i} - w_{i1}^{\text{поточне}}) \\ \text{Для Віку: } w_{11}^{\text{нове}} &= w_{11}^{\text{поточне}} + \eta (x_{11} - w_{11}^{\text{поточне}}) = 0,9 + 0,5 \times (0,8 - 0,9) = 0,85. \\ \text{Для Доходу: } w_{21}^{\text{нове}} &= w_{21}^{\text{поточне}} + \eta (x_{12} - w_{21}^{\text{поточне}}) = 0,8 + 0,5 \times (0,8 - 0,8) = 0,8. \end{aligned}$$

Дане налагоджування дозволить нейрону 1 у подальшому більш успішно захоплювати записи з інформацією про літніх людей з високим доходом.

Початкові значення ваг нейронів:

w_{11}	w_{21}	w_{12}	w_{22}	w_{13}	w_{23}	w_{14}	w_{24}
0,9	0,8	0,9	0,2	0,1	0,8	0,1	0,2

Набір записів вхідної вибірки:

№	x_{i1}	x_{i2}	Опис
1	$x_{11}=0,8$	$x_{12}=0,8$	Літня людина з високим доходом
2	$x_{21}=0,8$	$x_{22}=0,1$	Літня людина з низьким доходом
3	$x_{31}=0,2$	$x_{32}=0,8$	Молода людина з високим доходом
4	$x_{41}=0,1$	$x_{42}=0,2$	Молода людина з низьким доходом

Виконавши операції конкуренції та підстроювання для другого вхідного вектору $X_2=(0,8;0,1)$, отримуємо:

$D(W_1, X_2)$	$D(W_2, X_2)$	$D(W_3, X_2)$	$D(W_4, X_2)$	$w_{12} \text{ нове}$	$w_{22} \text{ нове}$
0,71	0,14	0,99	0,71	0,85	0,15

Переміг нейрон 2. Він відкриває кластер для захоплення літніх людей з малим доходом. Для третього і четвертого нейронів, відповідно, отримаємо такі нові значення ваг

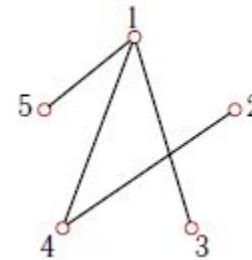
$w_{13} \text{ нове}$	$w_{23} \text{ нове}$	$w_{14} \text{ нове}$	$w_{24} \text{ нове}$
0,15	0,85	0,1	0,15

які будуть відповідати кластерам для молодих людей з високим доходом і молодих людей з низьким доходом.

Таким чином 4 вихідні нейрони представляють 4 різних кластера

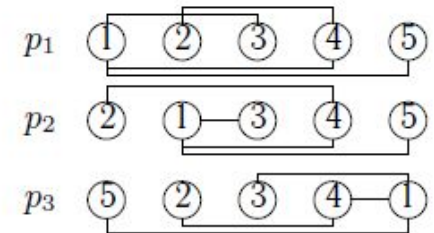
Кількість вихідних нейронів мережі Кохонена має відповідати кількості кластерів, які треба побудувати.

Знайдіть найкраще розташування вершин графу, за умов розміщення їх в один ряд, після трьох циклів роботи ГА при заданому початковому наборі хромосом. Якість розміщення оцінюється сумою довжини ребер графу. Єдиною операцією, що здійснюється на кожній ітерації роботи алгоритму є мутація, яка застосовується до кращої хромосоми покоління по розряду, який відповідає номеру ітерації і полягає у інверсії порядку розташування значень всіх генів хромосоми, розташованих за вибраним для мутації. Оцініть якість кожної з отриманих популяцій.

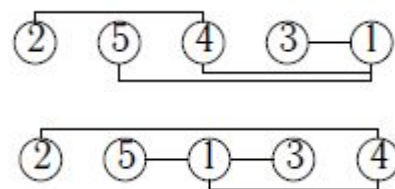


p_1	1	2	3	4	5
p_2	2	1	3	4	5
p_3	5	2	3	4	1

1. Розміщаємо хромосоми відповідно з генами (номерами вершин) хромосом.
2. Кількість горизонтальних відрізків між вершинами: $L_1=3+4+2+2=11$, $L_2=1+2+3+3=9$, $L_3=1+2+2+4=9$. Хромосома 2 є найменшою.



Піддаємо її мутації оператором інверсії по першому елементу. Тобто перша вершина залишається на своєму місці, а інші записуються у зворотному порядку: 25431. Довжина ребер: $L_4=1+2+3+2=8$. Отож, міняємо L_1 на L_4 і

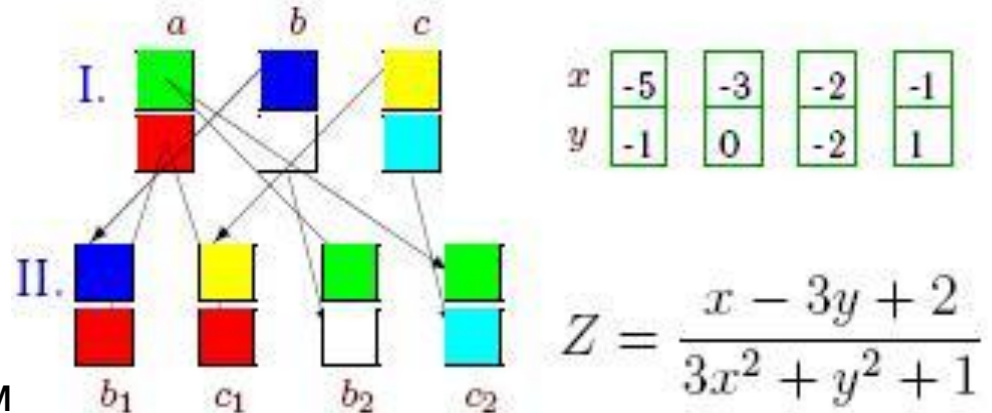


отримуємо другу популяцію. Краща хромосома після мутації набуде вигляду: 25134 з довжиною 8. Третя популяція:

p_1	2	5	4	3	1
p_2	2	1	3	4	5
p_3	2	5	1	3	4

p_1	2	5	4	3	1
p_2	2	1	3	4	5
p_3	5	2	3	4	1

Задано початкову популяцію з 4 хромосом, кожна з яких має по 2 гени x і y . Пристосованість хромосоми оцінюється функцією Z . При однакових Z перевагу має хромосома з більшим номером. На кожній ітерації найкраща хромосома a породжує 4 нові хромосоми b_1, c_1, b_2, c_2 , схрещенням з хромосомами b і c з більш низькими значеннями Z за схемою, наведеною на рисунку. Хромосома з найгіршою пристосованістю вилучається з популяції. Знайдіть показник найкращої пристосованості хромосоми в популяції і значення середньої пристосованості популяції після 3-х етапів еволюції.



Перше покоління (якість хромосом):

$$Z_{11}(-5, -1) = \frac{-5 + 3 \cdot 1 + 2}{3 \cdot (-5)^2 + (-1)^2 + 1} = 0 \quad Z_{12}(-3, 0) = \frac{-3 - 3 \cdot 0 + 2}{3 \cdot (-3)^2 + 1} = -\frac{1}{28} = -0.036$$

$$Z_{13}(-2, -2) = \frac{-2 + 3 \cdot 2 + 2}{3 \cdot (-2)^2 + (-2)^2 + 1} = \frac{6}{17} = 0.353 \quad Z_{14}(-1, 1) = \frac{-1 - 3 \cdot 1 + 2}{3 \cdot (-2)^2 + (1)^2 + 1} = -\frac{2}{5} = -0.4$$

$$I_1 = \sum_{i=1}^4 Z_{1i} = 0 - 0.036 + 0.353 - 0.4 = -0.083$$

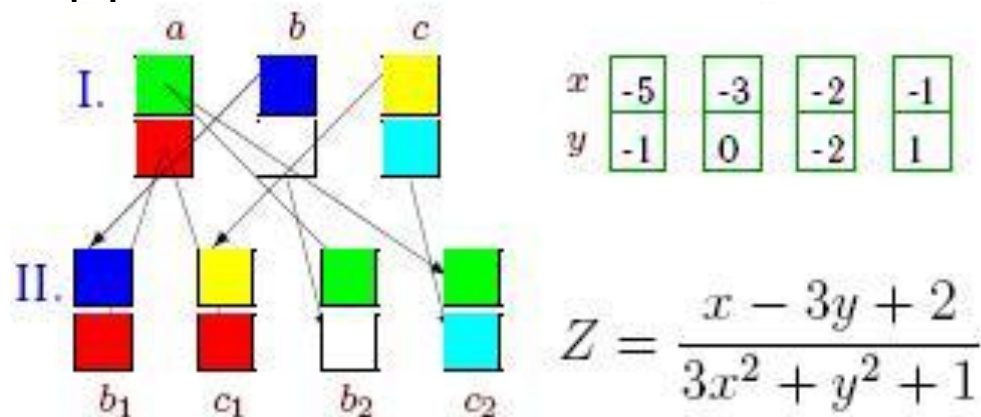
Якість популяції:

Вибираємо для наступного покоління три кращі хромосоми: Z_{13}, Z_{11}, Z_{12} . Наступне покоління буде породжено хромосомами:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -2 \\ -2 \end{pmatrix}, \begin{pmatrix} -5 \\ -1 \end{pmatrix}, \begin{pmatrix} -3 \\ 0 \end{pmatrix}$$

ГЕНЕТИЧНІ АЛГОРИТМИ (2)

15



Отримаємо з відібраних хромосом друге покоління $\begin{pmatrix} -5 \\ -2 \end{pmatrix}, \begin{pmatrix} -3 \\ -2 \end{pmatrix}, \begin{pmatrix} -2 \\ -1 \end{pmatrix}, \begin{pmatrix} -2 \\ 0 \end{pmatrix}$:

$$Z_{2,1}(-5, -2) = \frac{-5 + 3 \cdot 2 + 2}{3 \cdot (-5)^2 + (-2)^2 + 1} = 0.0375 \quad Z_{2,2}(-3, -2) = \frac{-3 + 3 \cdot 2 + 2}{3 \cdot (-3)^2 + (-2)^2 + 1} = 0.156$$

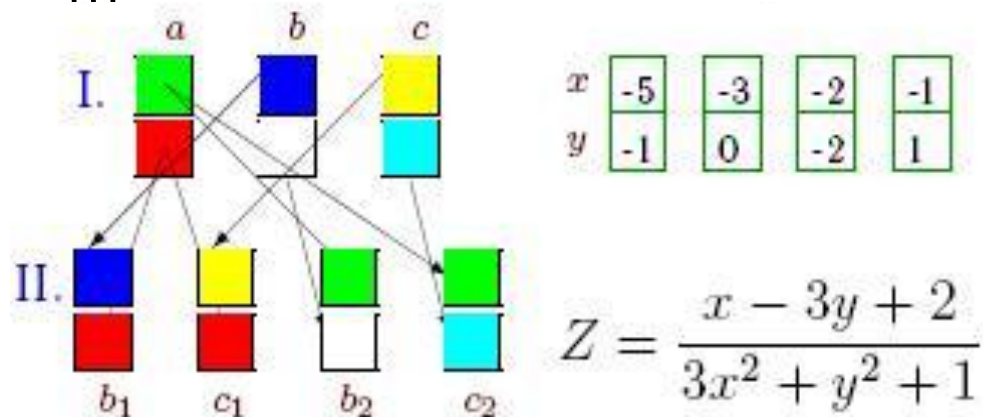
$$Z_{2,3}(-2, -1) = \frac{-2 + 3 \cdot 1 + 2}{3 \cdot (-2)^2 + (-1)^2 + 1} = 0.214 \quad Z_{2,4}(-2, 0) = \frac{-2 + 2}{3 \cdot (-2)^2 + 1} = 0$$

Якість другої популяції: $I_2 = \sum_4^{i=1} Z_{2,i} = 0.375 + 0.156 + 0.214 + 0 = 0.408$

Третє покоління буде породжене хромосомами: $\begin{pmatrix} -2 \\ -1 \end{pmatrix}, \begin{pmatrix} -3 \\ -2 \end{pmatrix}, \begin{pmatrix} -5 \\ -2 \end{pmatrix}$

ГЕНЕТИЧНІ АЛГОРИТМИ (3)

16



Отримаємо з відібраних хромосом третє покоління $\begin{pmatrix} -3 \\ -1 \end{pmatrix}, \begin{pmatrix} -5 \\ -1 \end{pmatrix}, \begin{pmatrix} -2 \\ -2 \end{pmatrix}, \begin{pmatrix} -2 \\ -2 \end{pmatrix}$:

$$Z_{3,1}(-3, -1) = \frac{-3 + 3 \cdot 1 + 2}{3 \cdot (-3)^2 + (-1)^2 + 1} = 0.069$$

$$Z_{3,2}(-5, -1) = \frac{-5 + 3 \cdot 1 + 2}{3 \cdot (-5)^2 + (-1)^2 + 1} = 0$$

$$Z_{3,3}(-2, -2) = \frac{-2 + 3 \cdot 2 + 2}{3 \cdot (-2)^2 + (-2)^2 + 1} = 0.353$$

$$Z_{3,4}(-2, -2) = \frac{-2 + 3 \cdot 2 + 2}{3 \cdot (-2)^2 + (-2)^2 + 1} = 0.353$$

Якість третьої популяції: $I_3 = \sum_{i=1}^4 Z_{3,i} = 0.069 + 0 + 0.353 + 0.353 = 0.775$

ГЕНЕТИЧНІ АЛГОРИТМИ (4)

17

Розглянемо пошук рішень діофантова рівняння (тільки цілі рішення) $a + 3b + 5c = 12$ (a, b, c – цілі) за допомогою ГА (для однієї генерації) при розмірі популяції $P=5$.

Хромосома, як можливе рішення, подається трьома числами (a, b і c) з діапазону $[1, 12]$
Хромосоми першої популяції вибираються випадковим чином.

Хромосома	(a,b,c)	Пристаєваність	1/n	Ймовір.	Батько	Мати
1	(3,6,2)	31-12=19	0,0526	29,9%	1	5
2	(1,12,8)	77-12=65	0,01538	8,8%	3	2
3	(2,9,0)	29-12=17	0,05882	33,5%	3	1
4	(11,5,9)	71-12=69	0,01449	8,2%	5	3
5	(8,1,6)	41-12=29	0,03448	19,6%	1	4
		Середнє: 39,8	$\sum = 0,17577$			

Відбір пар хромосом для схрещування здійснюємо на основі пропорційної селекції. Оскільки хромосома має три гени, можливі лише чотири види схрещування (розрив після першого або другого гену, та обмін лівими або правими частинами).

Зхрещування	Батько	Мати	Нащадок	Пристаєваність	Відбір
1л	(3,6,2)	(8,1,6)	(3,1,6)	36-12=24	(2,9,0)
2л	(2,9,0)	(1,12,8)	(2,9,8); (2,1,8)	69-12=57; 45-12=33	(3,9,0)
1п	(2,9,0)	(3,6,2)	(3,9,0)	30-12=18	(3,6,2)
2п	(8,1,6)	(2,9,0)	(2,9,6); (7,9,6)	59-12=47; 64-12=52	(3,1,6)
1л	(3,6,2)	(11,5,9)	(3,5,9)	63-12=51	(8,1,6)
				Середнє: 35,6	Середнє: 21,4

Здійснимо дві мутації, які полягають у випадковій заміні одного числа в хромосомі. Середня пристосованість нащадків вище ніж у батьків! Відберемо 5 кращих нащадків. Це знов можна зробити рулеткою, але для економії часу просто відберемо п'ять найбільш пристосованих хромосом (стовпчик 6). Бачимо, що середня пристосованість ще зросла і на наступному кроці еволюції є усі шанси отримати ще кращі хромосоми.