

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Київський національний університет будівництва і архітектури

Методичні вказівки
до проведення лабораторних занять
з дисципліни «Економіко-математичне моделювання»
для здобувачів першого (бакалаврського) рівня вищої освіти
за спеціальністю 075 «Маркетинг»

д.т.н., проф. Терейковська Л.О.

Практична робота №1

Визначення оптимального плану задачі лінійного програмування засобами оптимізації MS Excel

Мета: отримати практичні навички використання інструментарію аналізу даних середовища Excel для вирішення оптимізаційних задач.

Теоретичні відомості: лекція №1, 2.

Постановка задачі. Фабрика випускає два види фарб для зовнішніх (З) та внутрішніх (В) робіт, виробництво яких передбачає використання двох продуктів – А та Б. Максимально можливі запаси продуктів А та Б на добу складають 6 та 8 тон відповідно. Оптові ціни фарб З та В складають 3000 грн та 2000 грн. Витрати продуктів А та Б на 1 тону наведені в таблиці:

Витрати (продукт)	Витрати продуктів на 1 тону фарби		Максимально можливий запас на добу
	Фарба З	Фарба В	
А	1	2	6
Б	2	1	8

Крім цього встановлено, що:

- Різниця між попитом на фарбу В і попитом на фарбу З не більше 1 тони на добу.
- Попит на фарбу В не перевищує 2 тони на добу.

Необхідно визначити, який обсяг фарб кожного виду повинна виробляти фабрика, щоб отримати максимальний прибуток.

Хід роботи:

- Відповідно теоретичних відомостей, математична модель даної задачі має наступний вигляд: максимізувати функцію Z ,
 $Z = 3000 \cdot X_3 + 2000 \cdot X_В$ Обмеження моделі:

$X_3 + 2X_В$	≤ 6
$2X_3 + X_В$	≤ 8
$X_В - X_3$	≤ 1
$X_В$	≤ 2
$X_В$	≥ 0
X_3	≥ 0

- Запустити середовище Excel та створити файл pr_1.xlsx.
- Змінні моделі X_3 та $X_В$ визначено в комірках А3 та В3 відповідно. Цільова функція визначена в комірці D4, що показано на рис.1.

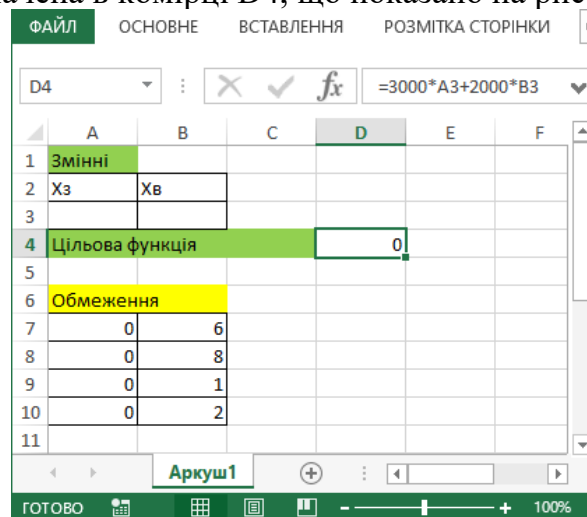


Рис.1

4. Обмеження визначені A7:A10 та B7:B10, що ілюструється на рис 2-5 (див. рядок формул).

A7						
	A	B	C	D	E	F
1	Змінні					
2	Хз	Хв				
3						
4	Цільова функція			0		
5						
6	Обмеження					
7	0	6				
8	0	8				
9	0	1				
10	0	2				

Рис.2

A8						
	A	B	C	D	E	F
1	Змінні					
2	Хз	Хв				
3						
4	Цільова функція			0		
5						
6	Обмеження					
7	0	6				
8	0	8				
9	0	1				
10	0	2				

Рис.3

A9						
	A	B	C	D	E	F
1	Змінні					
2	Хз	Хв				
3						
4	Цільова функція			0		
5						
6	Обмеження					
7	0	6				
8	0	8				
9	0	1				
10	0	2				

Рис.4

A10		:	X	✓	f_x	=B3
	A	B	C	D	E	
1	Змінні					
2	Xз	Xв				
3						
4	Цільова функція			0		
5						
6	Обмеження					
7	0	6				
8	0	8				
9	0	1				
10	0	2				
11						

Рис.5

5. Викликаємо сервіс Розв'язувач (меню Дані-Розв'язувач). Необхідно визначити обмеження для моделі. Для цього Натискаємо кнопку «Додати»

Параметри розв'язувача

Оптимізувати цільову функцію: SDS4

Діа:
☒ Максимум
☐ Мінімум
☐ Значення: 0

Змінюючи клітинки змінних: SAS3:SB53

Підлягає обмеженням:

Додати

Змінити

Видалити

Скинути

Завантажити/зберегти

☒ Зробити необмежені змінні не від'ємними

Виберіть метод розв'язання: За методом зведеного градієнта

Параметри

Метод розв'язання

Для розв'язання гладких нелінійних задач виберіть розв'язувач нелінійних задач за методом зведеного градієнта. Для розв'язання лінійних завдань виберіть розв'язувач за симплекс-методом, для негладких завдань виберіть розвиваний розв'язувач.

Довідка

Розв'язати

Закрити

Рис.6

Визначаємо 1 групу обмежень, які пов'язані з тим, що змінні в економічній моделі мають бути невід'ємними числами:

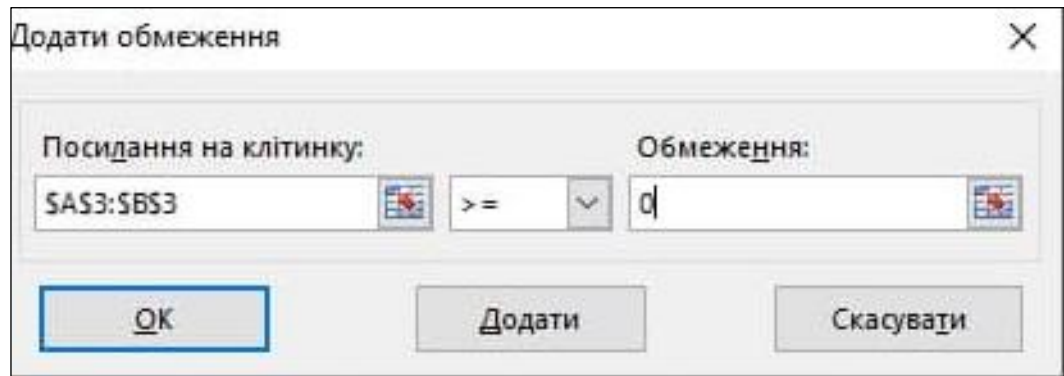


Рис.7

8. Визначаємо 2 групу обмежень, як показано на рис.8

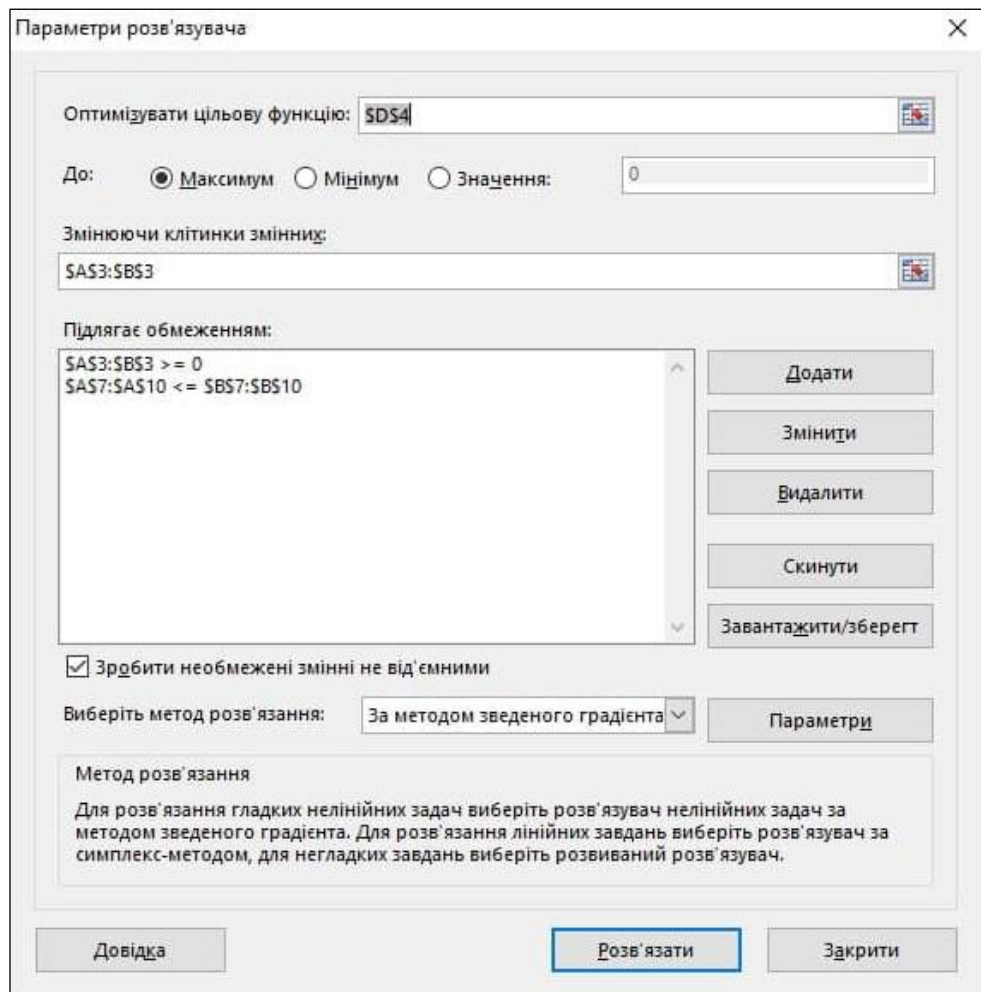


Рис.8

Самостійно визначити 3 та 4 обмеження моделі.

Натискаємо кнопку «Розв'язати»

9. Встановлюємо параметри розв'язання переключаючись на різні вкладки, як показано на рис. 9-11.

Параметри

Усі методи | За методом зведеного градієнта | Розвиваний розв'язувач

Точність обмеження: 0.000001

☒ Використовувати автоматичне масштабування

☐ Відображати результати ітерацій

Розв'язання з цілочисловими обмеженнями

☐ Ігнорувати цілочислові обмеження

Цілочислова оптимальність (%): 1

Ліміти розв'язання

Максимальний час (сек.):

Ітерації:

Розвинені та цілочислові обмеження:

Максимальна кількість підзадач:

Максимальна кількість допустимих розв'язань:

OK Скасувати

Рис.9

Параметри

Усі методи | За методом зведеного градієнта | Розвиваний розв'язувач

Конвергенція: 0.0001

Похідні

☒ Передні ☐ Центральні

Кілька початкових точок

☐ Використовувати кілька початкових точок

Розмір сукупності: 100

Випадкове початкове значення: 0

☒ Обов'язкові межі змінних

OK Скасувати

Рис.10

Параметри

Усі методи | За методом зведеного градієнта | Розвиваний розв'язувач

Конвергенція: 0.0001

Показник змінення: 0.075

Розмір сукупності: 100

Випадкове початкове значення: 0

Максимальний час без покращення: 30

☒ Обов'язкові межі змінних

Рис. 11

10. Натискаємо кнопку «Розв'язати»

Параметри розв'язувача

Оптимізувати цільову функцію: $SDS4$

До: ☒ Максимум ☐ Мінімум ☐ Значення: 0

Змінюючи клітинки змінних: $SAS3:SB53$

Підлягає обмеженням:

$SAS3:SB53 \geq 0$
 $SAS7:SAS10 \leq SB57:SB510$

☒ Зробити необмежені змінні не від'ємними

Виберіть метод розв'язання: За методом зведеного градієнта

Метод розв'язання

Для розв'язання гладких нелінійних задач виберіть розв'язувач нелінійних задач за методом зведеного градієнта. Для розв'язання лінійних завдань виберіть розв'язувач за симплекс-методом, для негладких завдань виберіть розвиваний розв'язувач.

Довідка Розв'язати Закрити

Рис. 12

11. Натискаємо кнопку «Зберегти сценарій»

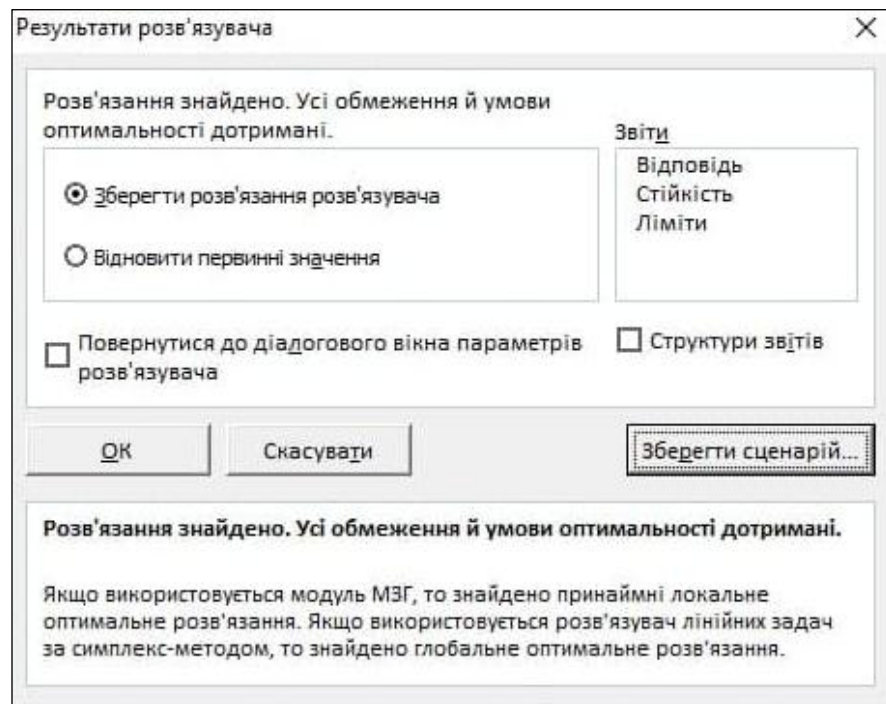


Рис. 13

12. В результаті пошуку оптимального рішення отримані дані у відповідні комірки.

	A	B	C	D	E
1	Змінні				
2	Хз	Хв			
3	3.333333	1.333333			
4	Цільова функція			12666.67	
5					
6	Обмеження				
7	6	6			
8	8	8			
9	-2	1			
10	1.333333	2			
11					

Рис. 14

У звіті до практичної роботи необхідно зробити висновки про отримані результати розв'язання поставленої задачі.

Практична робота №2

Побудова економіко-математичної моделі транспортної задачі

Мета: отримати практичні навички використання інструментарію аналізу даних середовища Excel для вирішення транспортних задач.

Теоретичні відомості: лекція №4.

Література: Економіко-математичне моделювання: Навчальний посібник / За ред. О. Т. Іващука. – Тернопіль: ТНЕУ «Економічна думка», 2020. – стор. 133-169.

Постановка задачі: Фірма має 4 фабрики, що розміщені у Чернігові, Львові, Харкові та Черкасах з виробничими можливостями 200 т, 150 т, 225 т та 175 т на добу, відповідно. Центри розподілення продукції фірми розміщені у Запоріжжі, Києві, Черкасах, Дніпрі та Львові (з потребами у продукції 150 т, 100 т, 200 т, 50 т, 250 т відповідно). Вартість перевезень одиниці продукції надано в таблиці:

Фабрика і центри	Київ	Черкаси	Дніпро	Львів	Запоріжжя
Чернігів	1,5	2	2,75	2,25	2,25
Львів	2,5	2	3	1	2,5
Харків	2	1,5	1,5	2,35	1,75
Черкаси	2	0,5	1,75	2,05	1,75

Завдання: спланувати обсяги перевезень так, щоб мінімізувати сумарні транспортні витрати.

Рішення: Побудуємо математичну модель для даної задачі. Невідомі – обсяги перевезень. Нехай X_{ij} – це обсяг перевезення з i -ої фабрики у j -ий центр розподілення.

Функція мети:

$$Z = \sum_{i=1}^4 \sum_{j=1}^5 (C_{ij} \times X_{ij}) \rightarrow \min$$

де C_{ij} - вартість перевезення одиниці продукції з i -ої фабрики у j -ий центр розподілення.

Обмеження на невідомі моделі:

1. Обсяг перевезень не може бути від'ємним числом ($X_{ij} \geq 0$).
2. Передбачається, що дана математична модель збалансована (тобто обсяг виготовленої продукції дорівнює сумарному попиту на неї з тобто в моделі не враховані витрати, пов'язані із зберіганням на складах).

Представимо вхідні дані в таблиці Excel:

	A	B	C	D	E	F	G
1	1,5	2	2,75	2,25	2,25	Чернігів	
2	2,5	2	3	1	2,5	Львів	
3	2	1,5	1,5	2,95	1,75	Харків	
4	2	0,5	1,75	2,05	1,75	Черкаси	
5	Київ	Черкаси	Дніпро	Львів	Запоріжжя		
6							200
7							150
8							225
9							175
10							
11	100	200	50	250	150		
12							

Рис. 1

1. В діапазоні A1:E4 – розміщено вартості перевезення (C_{ij}).
2. В діапазоні A6:E9 – змінні моделі (X_{ij} – обсяг перевезень).

- В діапазоні G6:G9 – обсяги виробництва (пропозиція) по фабриках - задані за умовою задачі).
- В діапазоні A11:E11 - потреби в продукції по різних містах (задані за умовою задачі).

	A	B	C	D	E	F	G	H
1	1,5	2	2,75	2,25	2,25	Чернігів		
2	2,5	2	3	1	2,5	Львів		
3	2	1,5	1,5	2,95	1,75	Харків		
4	2	0,5	1,75	2,05	1,75	Черкаси		
5	Київ	Черкаси	Дніпро	Львів	Запоріжжя			
6						0,00	200	Чернігів
7						0,00	150	Львів
8						0,00	225	Харків
9						0,00	175	Черкаси
10	0,00	0,00	0,00	0,00	0,00	0,00		
11	100	200	50	250	150			
12								

Рис. 2

В комірці A10 - підрахунок сумарного обсягу продукту (див. рядок формул), що звозиться у центр розподілу (в даному випадку у Київ). Розповсюджуємо цю формулу до E10.

- В комірку F6 вводимо F6=СУММ(A6:E6), розповсюджуємо цю формулу до F9

	A	B	C	D	E	F	G	H
1	1,5	2	2,75	2,25	2,25	Чернігів		
2	2,5	2	3	1	2,5	Львів		
3	2	1,5	1,5	2,95	1,75	Харків		
4	2	0,5	1,75	2,05	1,75	Черкаси		
5	Київ	Черкаси	Дніпро	Львів	Запоріжжя			
6						0,00	200	Чернігів
7						0,00	150	Львів
8						0,00	225	Харків
9						0,00	175	Черкаси
10	0,00	0,00	0,00	0,00	0,00	0,00		
11	100	200	50	250	150			
12								

Рис. 3

- Після цього визначаємо функцію мети (визначимо функцію мети у комірці F10) - див. рядок формул.

Параметри розв'язувача

Оптимізувати цільову функцію: 

До: ☐ Максимум ☒ Мінімум ☐ Значення:

Змінюючи клітинки змінних: 

Підлягає обмеженням:

☒ Зробити необмежені змінні не від'ємними

Виберіть метод розв'язання: 

Метод розв'язання

Для розв'язання гладких нелінійних задач виберіть розв'язувач нелінійних задач за методом зведеного градієнта. Для розв'язання лінійних завдань виберіть розв'язувач за симплекс-методом, для негладких завдань виберіть розвиваний розв'язувач.

Рис. 6

7. Представити скриншоти розв'язання задачі та зробити висновки по роботі.

Практична робота №3

Використання моделі регресійного аналізу для прогнозування

Мета: отримати практичні навички використання моделі регресійного аналізу для прогнозування.

Теоретичні відомості: лекція №9.

Література: Економіко-математичне моделювання: Навчальний посібник / За ред. О. Т. Іващука. – Тернопіль: ТНЕУ «Економічна думка», 2020. – стор. 415-465.

Завдання. Надано статистичні дані про одноденний середньодушовий прожитковий мінімум працездатної особи та її середньоденну зарплату (табл. 2) [2].

Таблиця 2 – Статистичні дані про одноденний середньодушовий прожитковий мінімум працездатної особи та її середньоденну зарплату

ID регіону	Середньодушовий прожитковий мінімум працездатної особи за день	Середньоденна зарплата
1	75	133
2	78	125
3	81	129
4	93	153
5	86	140
6	77	135
7	85	135
8	77	132
9	89	161
10	95	159
11	72	120
12	115	160

1. Побудувати лінійне рівняння парної регресії у від x .
2. Розрахувати лінійний коефіцієнт парної кореляції і середню помилку апроксимації.
3. Оцінити статистичну значущість параметрів регресії і кореляції за допомогою критеріїв Фішера і Стьюдента.
4. Виконати прогноз заробітної плати при прогнозному значенні середньодушового прожиткового мінімуму, що становить 107% від середнього рівня.
5. Оцінити точність прогнозу, розрахувавши помилку прогнозу і його довірчий інтервал.
6. На одному графіку побудувати вихідні дані і теоретичну пряму.

Розв'язання. Обчислимо характеристики випадкових величин X і Y (вибіркове середнє і вибіркове середнє квадратичне відхилення).

1. Для розрахунку параметрів рівняння лінійної регресії побудуємо розрахункову таблицю (табл. 3).

Таблиця 3 – Розрахункові дані для визначення параметрів рівняння лінійної регресії

№	x_i	y_i	$x_i y_i$	x_i^2	y_i^2	$\hat{y}_i$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$	$\frac{ y_i - \hat{y}_i }{y_i} \times 100$
1	75	133	9975	5625	17689	129.808	3.192	10.189	2.400
2	78	125	9750	6084	15625	132.844	-7.844	61.528	6.275
3	81	129	10449	6561	16641	135.88	-6.88	47.334	5.333
4	93	153	14229	8649	23409	148.024	4.976	24.761	3.252
5	86	140	12040	7396	19600	140.94	-0.94	0.884	0.671
6	77	135	10395	5929	18225	131.832	3.168	10.036	2.347
7	85	135	11475	7225	18225	139.928	-4.928	24.285	3.650
8	77	132	10164	5929	17424	131.832	0.168	0.028	0.127
9	89	161	14329	7921	25921	143.976	17.024	289.817	10.574
10	95	159	15105	9025	25281	150.048	8.952	80.138	5.630
11	72	120	8640	5184	14400	126.772	-6.772	45.860	5.643
12	115	160	18400	13225	25600	170.288	-10.288	105.843	6.430
Усього:	1023	1682	144951	88753	238040			700.703	52.334
Середнє значення	85.250	140.167	12079.250	7396.083	19836.667				4.361
σ	11.337	13.783							
σ^2	128.521	189.972							

$$a_1 = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2} = \frac{12079,5 - 140,167 \cdot 85,25}{7396,083 - 85,25^2} = 1,012.$$

$$a_0 = \bar{y} - a_1 \bar{x} = 140,167 - 1,012 \cdot 85,25 = 53,894.$$

Отже, рівняння лінійної регресії набуде вигляду:

$$y = 53,894 + 1,012x.$$

Це означає, що за зростання середньодушового прожиткового мінімуму на 1 грн середньоденна зарплата зростатиме в середньому на 1,012 грн.

2. З використанням коефіцієнта кореляції оцінимо щільність лінійного зв'язку

$$r_{xy} = a_1 \frac{\sigma_x}{\sigma_y} = 1,012 \cdot \frac{11,337}{13,783} = 0,832.$$

Коефіцієнт детермінації:

$$r_{xy}^2 = 0,832^2 = 0,692.$$

Отже, 69,2% варіації зарплати (y) обумовлюється варіацією фактора x – середньодушового прожиткового мінімуму.

Шляхом обчислення середньої помилки апроксимації, оцінимо якість побудованої моделі:

$$\bar{A} = \frac{1}{n} = \sum \frac{|y_i - \hat{y}_i|}{|y_i|} \cdot 100 = \frac{52,334}{12} = 4,361\%.$$

Якість побудованої моделі оцінюються як гарна, оскільки середня помилка апроксимації не перевищує 8–10%.

3. Розрахуємо F -критерій:

$$F_{\text{факт}} = \frac{r_{xy}^2}{1 - r_{xy}^2} (n-2) = \frac{0,692}{1 - 0,692} = 22,5.$$

За таблицею F -розподілу Фішера-Снедекора, при рівні значущості $\alpha = 0,05$ і кількості ступенів свободи $k_1=1$ і $k_2=12-2=10$, критичне значення становитиме:

$$F_{\text{таб}} = 4,96; \quad F_{\text{факт}} > F_{\text{таб}}.$$

H_0 – гіпотеза про статистичну незначимість рівняння регресії відхиляється.

Оцінення статистичної значущості параметрів регресії проведемо за допомогою t -статистики Стьюдента і шляхом довірчого інтервалу кожного з показників.

Висуваємо гіпотезу H_0 про статистичну незначимість відмінності показників від нуля:

$$a = b = r_{xy} = 0.$$

$t_{\text{табл}}$ для кількості ступенів свободи $df = n - 2 = 12 - 2 = 10$ і $\alpha = 0,05$ становитиме 2,23.

Визначимо випадкові помилки $m_a, m_b, m_{r_{xy}}$:

$$S_{\text{ост}} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}} = \sqrt{\frac{700,703}{12-2}} = 8,371;$$

$$m_{a_0} = \frac{S_{\text{ост}}}{n} \cdot \sqrt{\frac{\sum x^2}{\sum x^2 - \bar{x}^2}} = 8,371 \cdot \frac{\sqrt{7396,083}}{12 \cdot 11,377} = 5,292;$$

$$m_{a_1} = \frac{S_{\text{ост}}}{\sqrt{n(\bar{x}^2 - \bar{x}^2)}} = \frac{8,371}{11,377 \cdot \sqrt{12}} = 0,213;$$

$$m_{r_{xy}} = \cdot \sqrt{\frac{1-r_{xy}^2}{n-2}} = \sqrt{\frac{1-0,692}{12-2}} = 0,176.$$

Тоді:

$$t_{a_0} = \frac{a_0}{m_{a_0}} = \frac{53,894}{5,292} = 10,184;$$

$$|t_{a_0}| = 10,184 > t_{\text{табл}} = 2,23.$$

Фактичне значення є вищим за табличне значення t -статистики. Нульова гіпотеза відхиляється, оскільки a_0 не випадково відрізняється від нуля, а статистично значимо.

$$t_{a_1} = \frac{a_1}{m_{a_1}} = \frac{1,012}{0,213} = 4,751;$$

$$|t_{a_1}| = 4,751 > t_{\text{табл}} = 2,23.$$

Фактичне значення є вищим за табличне значення t -статистики. Нульова гіпотеза відхиляється, оскільки a_1 не випадково відрізняється від нуля, а статистично значимо.

$$t_{r_{xy}} = \frac{r_{xy}}{m_{r_{xy}}} = \frac{0,832}{0,176} = 4,727;$$

$$|t_{r_{xy}}| = 4,727 > t_{\text{табл}} = 2,23.$$

Фактичне значення є вищим за табличне значення t -статистики. Нульова гіпотеза відхиляється, оскільки $|t_{r_{xy}}|$ не випадково відрізняється від нуля, а статистично значимо.

Розрахуємо довірчі інтервали для параметрів регресії a_0 і a_1 . Для цього визначимо граничну помилку для кожного показника:

$$\Delta_{a_0} = t_{\text{табл}} \cdot m_{a_0} = 2,23 \cdot 5,392 = 12,024;$$

$$\Delta_{a_1} = t_{\text{табл}} \cdot m_{a_1} = 2,23 \cdot 0,213 = 0,475.$$

Довірчі інтервали:

$$\gamma_{a_0} = a_0 \pm \Delta_{a_0} = 53,894 \pm 12,024 \text{ або } 41,87 < a_0 < 65,918;$$

$$\gamma_{a_1} = a_1 \pm \Delta_{a_1} = 1,012 \pm 0,475 \text{ або } 0,537 < a_1 < 1,487.$$

4. Отримані оцінки рівняння регресії дозволяють використовувати його для прогнозу. Якщо прогнозне значення прожиткового мінімуму становитиме:

$$x_p = \bar{x} \cdot 1,07 = 85,25 \cdot 1,07 = 91,218 \text{ грн},$$

то прогнозне значення середньоденної зарплати буде:

$$y_p = 53,894 + 1,012 \cdot 91,218 = 146,207 \text{ грн.}$$

5. Помилка прогнозу становитиме:

$$m_p = S_{\text{ост}} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n\sigma_x^2}} = 9,521 \cdot \sqrt{1 + \frac{1}{12} + \frac{(91,218 - 85,25)^2}{12 \cdot 128,521}} = 10,015 \text{ грн.}$$

Гранична помилка прогнозу, яка не буде перевищена у 95% випадків, становитиме:

$$\Delta = t_{\text{набл}} \cdot m_p = 2,23 \cdot 10,015 = 22,334 \text{ грн.}$$

Довірчий інтервал прогнозу:

$$(146,207 - 22,334; 146,207 + 22,334) = (123,873; 168,541) \text{ грн.}$$

6. Побудуємо вихідні дані і теоретичну пряму

Практична робота №4

Використання мережі Кохонена для кластеризації даних

Мета: отримати навички використання мережі Кохонена для кластеризації даних.

Теоретичні відомості: лекція №15.

Література: Ковальчук О. Я. Математичне моделювання та прогнозування. : навчальний підручник. Тернопіль : ТНЕУ, 2019. – 412 с.

Завдання. Для кластеризації клієнтів банку за двома параметрами: віком та доходами, кожен з яких має по два значення. Потрібно здійснити навчання мережі Кохонена з використанням наданої банком навчальної вибірки (табл. 10) [3].

Таблиця 10 – Навчальна вибірка для кластеризації

№	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

Розв'язання. Вихідний шар мережі Кохонена буде мати 4 нейрони (рис. 12), оскільки їх кількість має дорівнювати кількості створюваних кластерів.

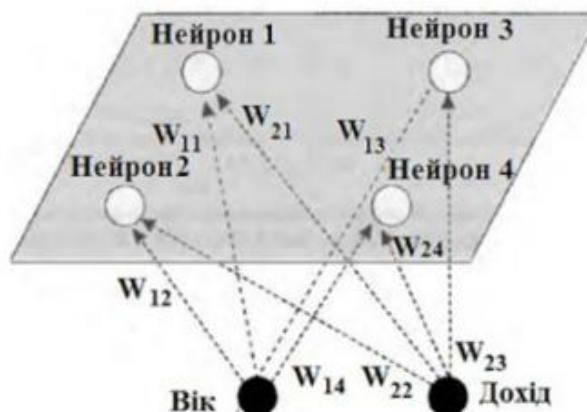


Рисунок 12 – Структура мережі Кохонена для кластеризації клієнтів за двома значеннями кожного з параметрів: «Вік» і «Доход»

На етапі ініціалізації, через надто малий розмір мережі, встановимо радіус навчання $R=0$. Таким чином, можливість налаштування вагових коефіцієнтів буде надано лише нейрону-переможцю. Коефіцієнт швидкості навчання встановимо на рівень $\eta = 0,5$ і згенеруємо випадковим чином початкові ваги нейронів (табл. 11).

Таблиця 11 – Початкові ваги 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

Ітерація 1. На етапі збудження подамо на перший нейрон перший вектор впливу з навчальної вибірки $X_1 (0,8; 0,8)$.

На етапі конкуренції обчислимо евклідову відстань між вхідним вектором X_1 і векторами ваг усіх чотирьох нейронів вихідного шару:

$$\begin{aligned}\text{Нейрон 1: } D(W_1, X_1) &= \sqrt{(w_{11} - x_{11})^2 + (w_{21} - x_{12})^2} = \\ &= \sqrt{(0,9 - 0,8)^2 + (0,8 - 0,8)^2} = 0,1.\end{aligned}$$

$$\begin{aligned}\text{Нейрон 2: } D(W_2, X_1) &= \sqrt{(w_{12} - x_{11})^2 + (w_{22} - x_{12})^2} = \\ &= \sqrt{(0,9 - 0,8)^2 + (0,2 - 0,8)^2} = 0,61.\end{aligned}$$

$$\begin{aligned}\text{Нейрон 3: } D(W_3, X_1) &= \sqrt{(w_{13} - x_{11})^2 + (w_{23} - x_{12})^2} = \\ &= \sqrt{(0,1 - 0,8)^2 + (0,8 - 0,8)^2} = 0,7.\end{aligned}$$

$$\begin{aligned}\text{Нейрон 4: } D(W_4, X_1) &= \sqrt{(w_{14} - x_{11})^2 + (w_{24} - x_{12})^2} = \\ &= \sqrt{(0,1 - 0,8)^2 + (0,2 - 0,8)^2} = 0,92.\end{aligned}$$

Найменшу відстань між вектором власних ваг $W_1 = (0,9; 0,8)$ і вхідним вектором навчальної вибірки $X_1 = (0,8; 0,8)$ має нейрон 1. Отож, саме він перемагає в конкуренції за право відкрити кластер, куди будуть вноситися літні люди з високим доходом.

Оскільки через малий розмір мережі радіус навчання має значення $R=0$, етап об'єднання, що мав би містити сусідні нейрони переможця, пропускається.

На етапі налаштування здійснимо налаштування ваг лише нейрона 1 (нейрона-переможця) згідно з формулою:

$$w_{ij,\text{нове}} = w_{ij,\text{поточне}} + \eta \cdot (x_{ni} - w_{ij,\text{поточне}}),$$

де $\eta = 0,5$;

$j=1$ для першого нейрона;

$n=1$ для першого запису.

Отже, для ознаки «Вік» отримуємо:

$$w_{11,\text{нове}} = w_{11,\text{поточне}} + 0,5 \cdot (x_{11} - w_{11,\text{поточне}}) = 0,9 + 0,5 \cdot (0,8 - 0,9) = 0,85.$$

Для ознаки «Дохід» отримуємо:

$$w_{21,\text{нове}} = w_{21,\text{поточне}} + 0,5 \cdot (x_{12} - w_{21,\text{поточне}}) = 0,8 + 0,5 \cdot (0,8 - 0,8) = 0,8.$$

Ітерація 2. Застосуємо дії, аналогічні діям ітерації 1, до другого вектора навчальної вибірки $X_2 = (0,8; 0,1)$.

$$\begin{aligned}\text{Нейрон 1: } D(W_1, X_2) &= \sqrt{(w_{11} - x_{21})^2 + (w_{21} - x_{22})^2} = \\ &= \sqrt{(0,85 - 0,8)^2 + (0,8 - 0,1)^2} = 0,71.\end{aligned}$$

$$\begin{aligned}\text{Нейрон 2: } D(W_2, X_2) &= \sqrt{(w_{12} - x_{21})^2 + (w_{22} - x_{22})^2} = \\ &= \sqrt{(0,9 - 0,8)^2 + (0,2 - 0,1)^2} = 0,14.\end{aligned}$$

$$\begin{aligned}\text{Нейрон 3: } D(W_3, X_2) &= \sqrt{(w_{13} - x_{21})^2 + (w_{23} - x_{22})^2} = \\ &= \sqrt{(0,1 - 0,8)^2 + (0,8 - 0,1)^2} = 0,99.\end{aligned}$$

$$\begin{aligned}\text{Нейрон 4: } D(W_4, X_2) &= \sqrt{(w_{14} - x_{21})^2 + (w_{24} - x_{22})^2} = \\ &= \sqrt{(0,1 - 0,8)^2 + (0,2 - 0,1)^2} = 0,71.\end{aligned}$$

Найближчим до вхідного вектора другого запису з навчальної вибірки $X_2 = (0,8; 0,1)$, порівняно з векторами ваг інших нейронів, виявився вектор ваг другого нейрона $W_2 = (0,9; 0,2)$. Отже, саме він і буде «засновником» другого кластера, що міститиме записи про літніх людей з низькими доходами.

Для нейрона 2 і другого запису навчальної вибірки мають місце значення $j = 2$ та $n = 2$, відповідно. Отже, налаштування вагових коефіцієнтів другого нейрона відбудеться за формулою:

$$w_{i2, \text{нове}} = w_{i2, \text{поточне}} + \eta \cdot (x_{2i} - w_{i2, \text{поточне}}).$$

Для ознаки «Вік» оновлене значення ваги становитиме:

$$w_{12, \text{нове}} = w_{12, \text{поточне}} + 0,5 \cdot (x_{21} - w_{12, \text{поточне}}) = 0,9 + 0,5 \cdot (0,8 - 0,9) = 0,85.$$

Ознака «Дохід» отримає таке оновлене значення ваги:

$$w_{22, \text{нове}} = w_{22, \text{поточне}} + 0,5 \cdot (x_{22} - w_{22, \text{поточне}}) = 0,2 + 0,5 \cdot (0,1 - 0,2) = 0,15.$$

Ітерація 3. Виконаємо аналогічні дії з третім вектором навчальної вибірки $X_3 = (0,2; 0,9)$.

$$\begin{aligned}\text{Нейрон 1: } D(W_1, X_3) &= \sqrt{(w_{11} - x_{31})^2 + (w_{21} - x_{32})^2} = \\ &= \sqrt{(0,85 - 0,2)^2 + (0,8 - 0,9)^2} = 0,66.\end{aligned}$$

$$\begin{aligned}\text{Нейрон 2: } D(W_2, X_3) &= \sqrt{(w_{12} - x_{31})^2 + (w_{22} - x_{32})^2} = \\ &= \sqrt{(0,85 - 0,2)^2 + (0,15 - 0,9)^2} = 0,99.\end{aligned}$$

$$\begin{aligned}\text{Нейрон 3: } D(W_3, X_3) &= \sqrt{(w_{13} - x_{31})^2 + (w_{23} - x_{32})^2} = \\ &= \sqrt{(0,1 - 0,2)^2 + (0,8 - 0,9)^2} = 0,14.\end{aligned}$$

$$\begin{aligned}\text{Нейрон 4: } D(W_4, X_3) &= \sqrt{(w_{14} - x_{31})^2 + (w_{24} - x_{32})^2} = \\ &= \sqrt{(0,1 - 0,2)^2 + (0,2 - 0,9)^2} = 0,71.\end{aligned}$$

Найближчим до вхідного вектора третього запису з навчальної вибірки $X_3 = (0,2; 0,9)$ виявився вектор ваг третього нейрона $W_3 = (0,1; 0,8)$. Таким чином, він стане «засновником» кластера для молодих людей з високими доходами.

Для нейрона 3 і третього запису навчальної вибірки мають місце значення $j = 3$ та $n = 3$, відповідно. Отже, налаштування вагових коефіцієнтів третього нейрона відбудеться за формулою:

$$w_{i3, \text{нове}} = w_{i3, \text{поточне}} + \eta \cdot (x_{3i} - w_{i3, \text{поточне}}),$$

Для ознаки «Вік» оновлене значення ваги становитиме:

$$w_{13, \text{нове}} = w_{13, \text{поточне}} + 0,5 \cdot (x_{31} - w_{13, \text{поточне}}) = 0,1 + 0,5 \cdot (0,2 - 0,1) = 0,15.$$

Ознака «Дохід» отримає таке оновлене значення ваги:

$$w_{23, \text{нове}} = w_{23, \text{поточне}} + 0,5 \cdot (x_{32} - w_{23, \text{поточне}}) = 0,8 + 0,5 \cdot (0,9 - 0,8) = 0,85.$$

Ітерація 4. Здійснимо навчання мережі Кохонена на четвертому записі з навчальної вибірки $X_4 = (0,1; 0,1)$.

$$\begin{aligned}\text{Нейрон 1: } D(W_1, X_4) &= \sqrt{(w_{11} - x_{41})^2 + (w_{21} - x_{42})^2} = \\ &= \sqrt{(0,85 - 0,1)^2 + (0,8 - 0,1)^2} = 1,05.\end{aligned}$$

$$\begin{aligned}\text{Нейрон 2: } D(W_2, X_4) &= \sqrt{(w_{12} - x_{41})^2 + (w_{22} - x_{42})^2} = \\ &= \sqrt{(0,85 - 0,1)^2 + (0,15 - 0,1)^2} = 0,75\end{aligned}$$

$$\begin{aligned}\text{Нейрон 3: } D(W_3, X_4) &= \sqrt{(w_{13} - x_{41})^2 + (w_{23} - x_{42})^2} = \\ &= \sqrt{(0,1 - 0,1)^2 + (0,8 - 0,1)^2} = 0,7.\end{aligned}$$

$$\begin{aligned}\text{Нейрон 4: } D(W_4, X_4) &= \sqrt{(w_{14} - x_{41})^2 + (w_{24} - x_{42})^2} = \\ &= \sqrt{(0,1 - 0,1)^2 + (0,2 - 0,1)^2} = 0,1.\end{aligned}$$

Найближчим до вхідного вектора четвертого запису з навчальної вибірки $X_4 = (0,1; 0,1)$ виявився вектор ваг четвертого нейрона $W_4 = (0,1; 0,2)$. Таким чином, він стає «засновником» кластера для молодих людей з низьким доходом.

Для нейрона 4 і четвертого запису навчальної вибірки мають місце значення $j = 4$ та $n = 4$, відповідно. Отже, налаштування вагових коефіцієнтів четвертого нейрона відбудеться за формулою:

$$w_{i4,\text{нове}} = w_{i4,\text{поточне}} + \eta \cdot (x_{4i} - w_{i4,\text{поточне}}),$$

Для ознаки «Вік» оновлене значення ваги не зміниться:

$$w_{14,\text{нове}} = w_{14,\text{поточне}} + 0,5 \cdot (x_{41} - w_{14,\text{поточне}}) = 0,1 + 0,5 \cdot (0,1 - 0,1) = 0,1.$$

Ознака «Дохід» отримає оновлене значення ваги:

$$w_{24,\text{нове}} = w_{24,\text{поточне}} + 0,5 \cdot (x_{42} - w_{24,\text{поточне}}) = 0,2 + 0,5 \cdot (0,1 - 0,2) = 0,15.$$

Отже, під час навчання мережі Кохонена чотири вихідні нейрони налаштовані на формування 4-х окремих кластерів, наведених у таблиці 12.

Таблиця 12 – Результати кластеризації

№ кластера	№ нейрона	Опис кластера
1	1	Літня людина з високим доходом
2	2	Літня людина з низьким доходом
3	3	Молода людина з високим доходом
4	4	Молода людина з низьким доходом

Практична робота №5

Вирішення практичних задач за допомогою багатошарового персептрону

Мета: отримати практичні навички використання нейронних мереж для вирішення маркетингових задач.

Теоретичні відомості: лекція №14.

Література: Ковальчук О. Я. Математичне моделювання та прогнозування : навчальний підручник. Тернопіль : ТНЕУ, 2019. – 412 с.

Постановка задачі: спрогнозувати обсяги виробництва фарби 1 та фарби 2. Статистика роботи фабрики з виробництва двох видів фарб представлена у файлі Viborka.xls:

	A	B	C	D	E
1	c1	c2	z	y1	y2
2	3000,00	2000,00	12650,00	3,33	1,33
3	2576,00	1950,00	10308,26	3,01	1,31
4	2311,00	1777,00	9123,78	2,71	1,61
5	3072,00	2002,00	12007,06	3,12	1,21
6	2890,00	1899,00	9789,48	2,52	1,32
7	2456,00	1950,00	10539,46	3,41	1,11
8	2900,00	1852,00	10456,68	2,75	1,34
9	3156,00	2130,00	13091,46	3,21	1,39
10	2963,00	2133,00	12130,00	3,05	1,45
11	3000,00	2100,00	9693,00	2,16	1,53
12	1999,00	1777,00	8820,17	3,31	1,24
13	3005,00	2007,00	12314,26	3,41	1,03
14	2999,00	1889,00	11830,50	3,00	1,50
15	3500,00	2400,00	12688,00	2,72	1,32
16	3200,00	2515,00	11719,95	2,46	1,53
17	3111,00	2111,00	10904,00	2,46	1,54
18	2973,00	2173,00	12218,50	3,05	1,45
19	3086,00	2072,00	12135,44	3,12	1,21
20	2596,00	2050,00	11018,66	3,21	1,31
21	3700,00	2423,00	14745,96	2,99	1,52
22	3700,00	2417,76	14833,00	2,99	1,42
23	3100,00	2050,00	14000,00	0,00	0,00

Рис.1

Інформація (рис.1) представлена за різні періоди діяльності фабрики:

c1 (ціна фарби 1, грн),

c2 (ціна фарби 2, грн),

z (прибуток чистий, тис грн),

y1 (обсяг виробництва фарби 1, т),

y2 (обсяг виробництва фарби 2, т).

З 2-го по 22-й рядок (сірий та жовтий колір) – це навчальна вибірка, при цьому c1, c2, z – вхідні параметри нейромережевої моделі. В останньому рядку (зелений колір) – тестовий приклад, для прогнозування значень y1 та y2 (вихідні параметри нейромережевої моделі).

Хід роботи:

1. Опрацювання лекційного матеріалу (лекція №12).
2. Опрацювання рекомендацій навч. посібника (Ковальчук О. Я. Математичне моделювання та прогнозування : навчальний підручник. Тернопіль : ТНЕУ, 2019).

- Визначення структурних параметрів нейронної мережі (БШП) та функції активації нейронів.

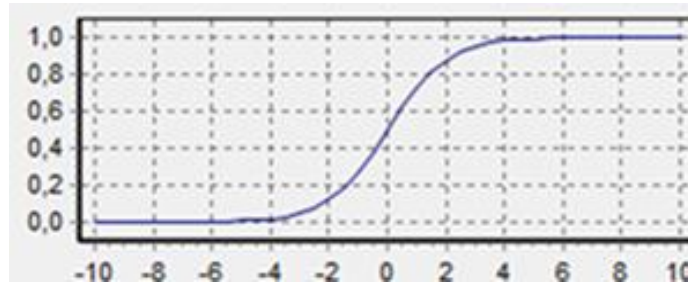


Рис.2 Графік сигмоїдальної функції активації

- Визначення алгоритму навчання нейромережі алгоритм зворотнього розповсюдження помилки (Back - Propagation)
- Встановлюємо обмеження по кількості епох навчання нейромережі.
- Визначення топології нейромережі (3х3х2)

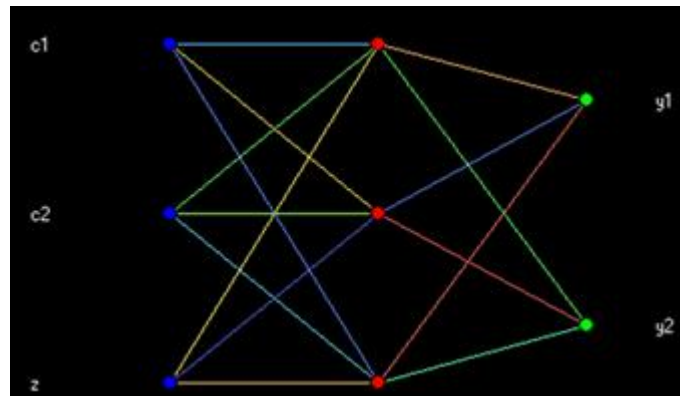


Рис. 3 Граф нейромережі

- Аналіз результату прогнозу, що виданий нейромережею.

c1	c2	z	y1	y2	y1_OUT	y2_OUT	y1_ERR	y2_ERR
3000	2000	12650	3.33	1.33	3.24224779627564	1.26339177273624	0.0006622276432506	0.00171160678184486
2576	1950	10308.26	3.01	1.31	3.03023730040082	1.35138465857092	3.5220571504643E-5	0.000660734526072062
2311	1777	9123.78	2.71	1.61	2.99782164991266	1.3294900514726	0.00712423372334635	0.0303560168291521
3072	2002	12007.06	3.12	1.21	3.05786744598152	1.35258657922168	0.000331993556028873	0.00784342138580348
2890	1899	9789.48	2.52	1.32	2.46597185825183	1.4404630982176	0.000251033281512928	0.00559830177546482
2456	1950	10539.46	3.41	1.11	3.18253664341485	1.28977232017741	0.00444952989645609	0.012467916786378
2900	1852	10456.68	2.75	1.34	2.77512925225655	1.39152627690341	5.43063199467968E-5	0.0010242495318572
3156	2130	13091.46	3.21	1.39	3.20398306912644	1.31639227238362	3.11344563060333E-6	0.00209023477676283
2963	2133	12130	3.05	1.45	3.12727906450391	1.35203461078067	0.000513588102149055	0.00370248735962883
3000	2100	9693	2.16	1.53	2.15047365665209	1.48365632362157	7.80447515779594E-6	0.000828570016692649
1999	1777	8820.17	3.31	1.24	3.17380743889783	1.25468716938217	0.00159513709888699	8.32193759733969E-5
3005	2007	12314.26	3.41	1.03	3.18168397805995	1.30467246782771	0.00448295128821799	0.0291057307135398
2999	1889	11830.5	3	1.5	3.1180618288594	1.31083593851064	0.00119869930888328	0.0138046534312533
3500	2400	12688	2.72	1.32	2.57813374313664	1.47057780448569	0.00173081026448181	0.00874722240798208
3200	2515	11719.95	2.46	1.53	2.51447227432612	1.48021561068791	0.000255177429697059	0.000956168905203452
3111	2111	10904	2.46	1.54	2.49889966508556	1.45539843493619	0.000130131658978579	0.0027612456352939
2973	2173	12218.5	3.05	1.45	3.12658271978297	1.3590003339837	0.000504374142753945	0.00319468354426085
3086	2072	12135.44	3.12	1.21	3.05136967944879	1.36636126765967	0.000405063673253708	0.00943206127238079
2596	2050	11018.66	3.21	1.31	3.16061684813322	1.32154022111558	0.000209724347769406	5.13779188289588E-5
3700	2423	14745.96	2.99	1.52	3.10466536339695	1.4040801271187	0.00113072174846466	0.00518398863038302
3700	2417.76	14833	2.99	1.42	3.12743595782563	1.39687571402583	0.00162437501565468	0.00020629319326523
3100	2050	14000	0	0	3.32773516820769	1.15490912231518	0.952332827351526	0.514563299335216

Рис. 4

Проаналізувати дані, які нейромережа видає в режимі розпізнавання (позначено жовтим кольором y1_OUT, y2_OUT, y1_ERR, y2_ERR), та зробити висновки у звіті до практичної роботи.

На рис. 5 представлена діаграма розсіювання, за допомогою якої можливо оцінити якість навчання нейронної мережі.

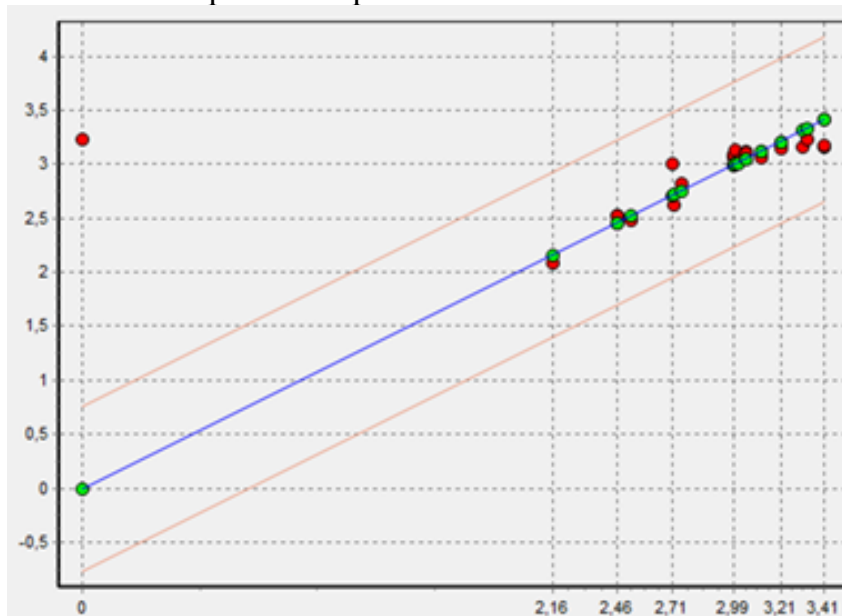


Рис. 5

Проаналізувати дані, що відображає діаграма розсіювання та зробити висновки.

Практична робота №6

Використання дерев рішень для класифікації даних

Мета: отримати практичні навички вирішення практичних задач за допомогою багатоваріантного перцептрону.

Теоретичні відомості: лекція №14.

Література: Ковальчук О. Я. Математичне моделювання та прогнозування : навчальний підручник. Тернопіль : ТНЕУ, 2019. – 412 с.

Завдання. Для проведення чергової маркетингової компанії відділ маркетингу проводить дослідження із визначення, які групи покупців найбільш схильні до придбання комп'ютера. Наявна інформація (табл. 4) щодо:

- віку (≤ 30 / $31 - 40$ / > 40);
- доходу (низький / середній / високий);
- кредитної історії (гарна / відмінна);
- статусу (студент / не студент);
- історії купівлі комп'ютера за результатами минулих маркетингових компаній.

Таблиця 4 – Начальна вибірка для побудови дерева рішень щодо схильності клієнта до 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
Студент	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
Придбання	ні	ні	так	так	так	ні	так	ні	так	так	так	так	так	ні

Ціль дослідження – за наявними характеристиками клієнта за допомогою алгоритму ID3 побудувати дерево рішень, що дозволить виявити ступінь зацікавленості клієнта у придбанні комп'ютера.

Розв'язання. При побудові дерева рішень для вибору атрибуту для розбиття скористуємося критерієм Gain приросту ентропії.

1. Розрахуємо ентропію кореневої вершини дерева рішень:

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

2. Розрахуємо приріст ентропії, який забезпечить розбиття дерева за кожним з наявних атрибутів:

а) атрибут «Вік» має три значення :

« ≤ 30 »: так – 2, ні – 3; « 31..40 »: так - 4, ні - 0, « > 40 »: так – 3, ні – 2.

Отримуємо:

$$\begin{aligned} \text{Entropy}(\text{Вік}) &= 5/14 (-2/5 \log(2/5) - 3/5 \log(3/5)) + 4/14 (0) + \\ &+ 5/14 (-3/5 \log(3/5) - 2/5 \log(2/5)) = 5/14(0.9709) + 0 + 5/14(0.9709) = 0.6935. \\ \text{Gain}(\text{Вік}) &= 0.94 - 0.6935 = 0.2465. \end{aligned}$$

б) атрибут «Дохід» має три значення:

«високий»: так – 2, ні – 2; «середній»: так – 4, ні – 2; «низький»: так – 3, ні – 1.

$$\begin{aligned} \text{Entropy(дохід)} &= 4/14(-2/4\log(2/4)-2/4\log(2/4)) + \\ &+ 6/14(-4/6\log(4/6)-2/6\log(2/6)) + 4/14(-3/4\log(3/4)-1/4\log(1/4)) = \\ &= 4/14(1) + 6/14(0.918) + 4/14(0.811) = 0.285714 + 0.393428 + 0.231714 = \\ &= 0.9108. \end{aligned}$$

$$\text{Gain(дохід)} = 0.94 - 0.9108 = 0.0292.$$

в) атрибут «Студент» має два значення:

«так»: так – 6, ні – 1; «ні»: так – 3, ні – 4.

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

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

в) атрибут «Кредитна історія» має два значення:

«гарна»: так – 6, ні – 2; «відмінна»: так – 3, ні – 2.

$$\begin{aligned} \text{Entropy(кредит)} &= 8/14(-6/8\log(6/8)-2/8\log(2/8)) + 6/14(-3/6\log(3/6) - \\ &-3/6\log(3/6)) = 8/14(0.8112) + 6/14(1) = 0.4635 + 0.4285 = 0.8920 \\ \text{Gain(Кредит)} &= 0.94 - 0.8920 = 0.048. \end{aligned}$$

Оскільки найбільше значення Gain має атрибут «Вік», починаємо побудову дерева рішень з його розбиття (рис. 3):

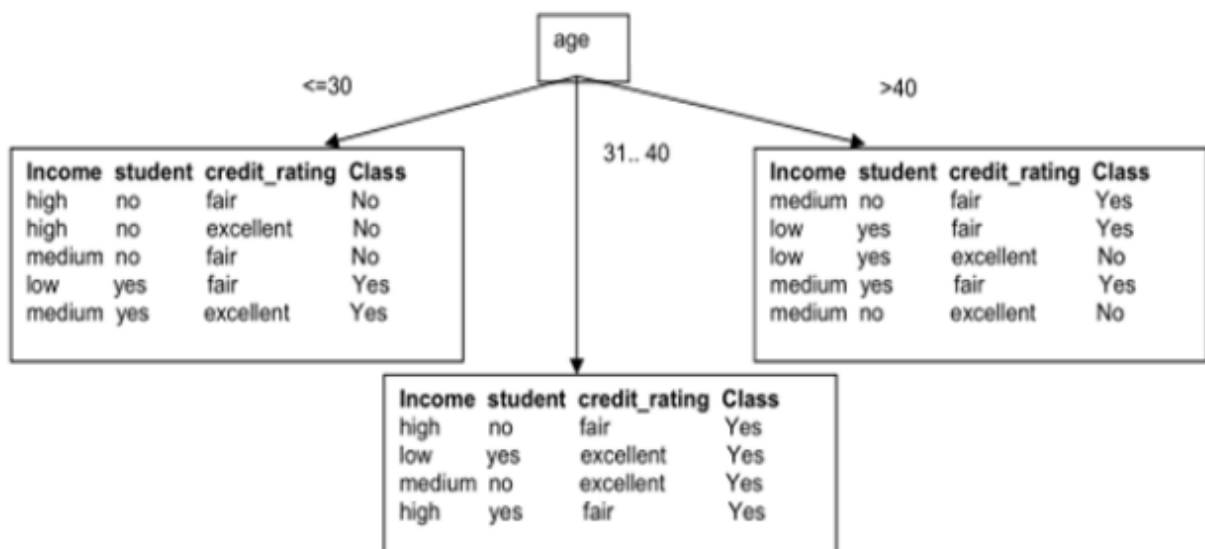


Рисунок 3 – Початок побудови дерева рішень з розбиття атрибуту «Вік»

Оскільки всі записи у гілці з віком «31..40» належать одному класу «так», замінимо отриманий блок на лист зі значенням «Придбати» = «так» (рис. 4).

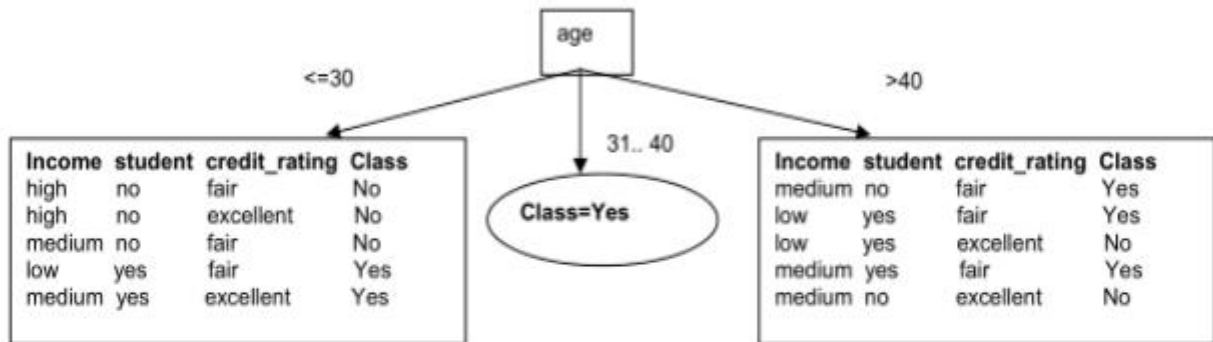


Рисунок 4 – Результат розбиття дерева рішень по атрибуту «Вік»

3. Розглянемо, за яким атрибутом потрібно розбити гілку «Вік» = «<=30».

Ентропія блока «<=30»:

$$I(S_{\text{ТАК}}, S_{\text{НІ}}) = I(2,3) = -2/5 \log_2 (2/5) - 3/5 \log_2 (3/5) = 0.97.$$

а) атрибут «Дохід» має три значення:

«високий»: так – 0, ні – 2; «середній»: так – 1, ні – 1; «низький»: так – 1, ні – 0.

$$\text{Entropy}(\text{Дохід}) = 2/5(0) + 2/5 (-1/2 \log_2 (1/2) - 1/2 \log_2 (1/2)) + 1/5 (0) = 2/5 (1) = 0,4.$$

$$\text{Gain}(\text{Дохід}) = 0,97 - 0,4 = 0,57.$$

б) атрибут «Студент» має два значення:

«так»: так – 2, ні – 0, «ні»: так – 0, ні – 3.

$$\text{Entropy}(\text{Студент}) = 2/5(0) + 3/5(0) = 0.$$

$$\text{Gain}(\text{Студент}) = 0,97 - 0 = 0,97.$$

Оскільки значення показника Gain для атрибуту «Студент» є максимальним, можна робити розбиття дерева за ним, без перевірки інших атрибутів (рис. 5).

Оскільки дві нових гілки є чистими класами, перетворимо їх на два листи з відповідними мітками (рис. 6).

Розглянемо гілку «>40», що залишилась.
Ентропія блоку «>40»:

$$I(S_{\text{ТАК}}, S_{\text{НІ}}) = I(3,2) = -3/5 \log_2 (3/5) - 2/5 \log_2 (2/5) = 0.97.$$

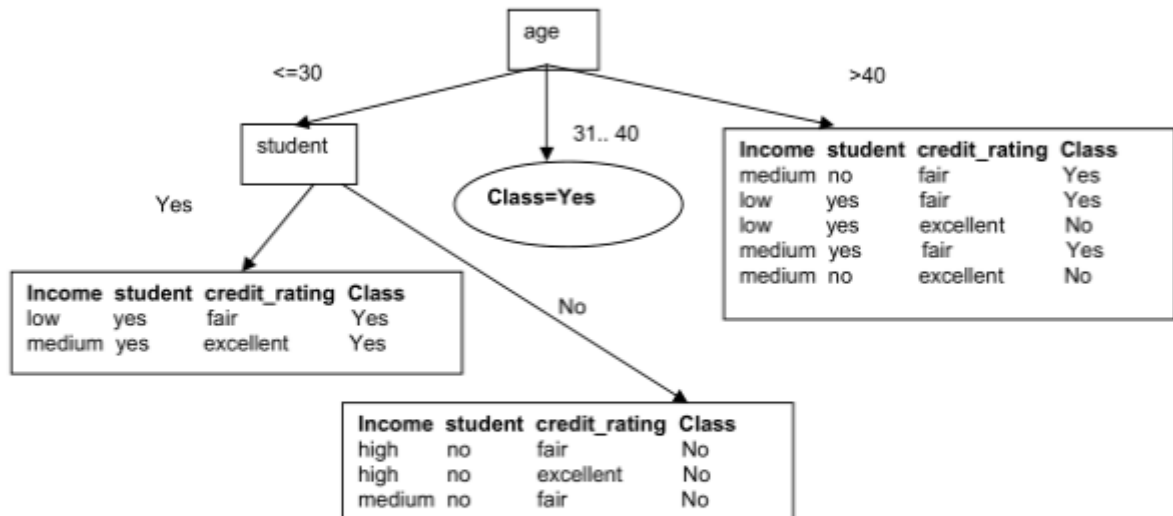


Рисунок 5 – Побудова дерева рішень розбиттям атрибуту «Студент»



Рисунок 6 – Результат розбиття дерева рішень за атрибутом «Студент»

а) атрибут «Дохід» має два значення:

«середній»: так – 2, ні – 1; «низький»: так – 1, ні – 1.

$$\text{Entropy(Дохід)} = 3/5(-2/3\log_2(2/3) - 1/3\log_2(1/3)) + 2/5(-1/2\log_2(1/2) - 1/2\log_2(1/2)) = 3/5(0.9182) + 2/5(1) = 0.55 + 0.4 = 0.95.$$

$$\text{Gain(Дохід)} = 0.97 - 0.95 = 0.02.$$

б) атрибут «Студент» має два значення:

«так»: так – 2, ні – 1; «ні»: так – 1, ні – 1.

$$\begin{aligned} \text{Entropy(Студент)} &= 3/5(-2/3\log(2/3) - 1/3\log(1/3)) + \\ &+ 2/5(-1/2\log(1/2) - 1/2\log(1/2)) = 0,95. \\ \text{Gain(Студент)} &= 0,97 - 0,95 = 0,02. \end{aligned}$$

в) атрибут «Кредитної історія» має два значення:

«гарна»: так – 3, ні – 0; «відмінна»: так – 0, ні – 2.

$$\text{Entropy(Кредит)} = 0.$$

$$\text{Gain(Кредит)} = 0,97 - 0 = 0,97.$$

Здійснимо розбиття за атрибутом «Кредит», яке надасть два чистих класи.

Побудоване дерево рішень набуде завершеного вигляду:

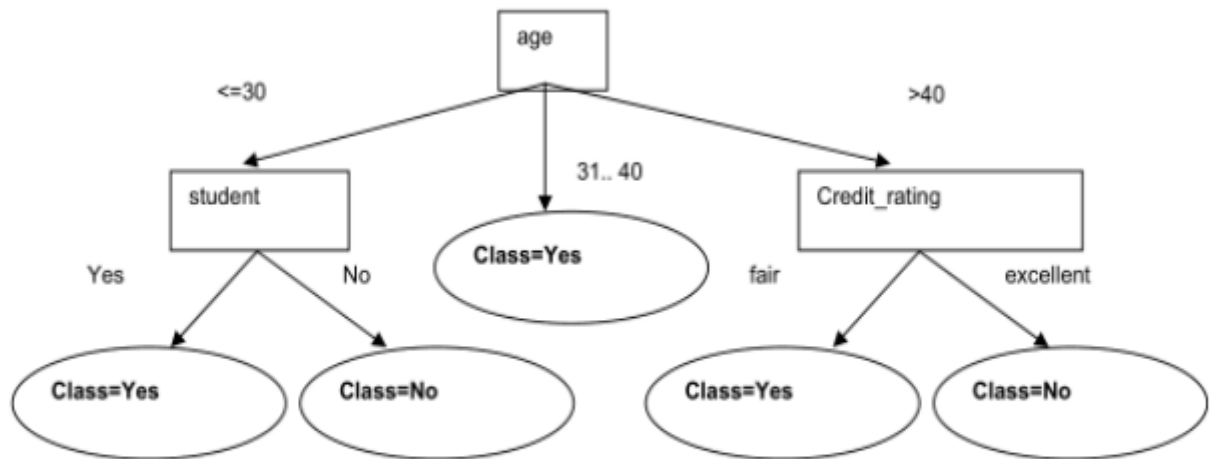


Рисунок 7 – Дерево рішень для класифікації клієнтів відносно їх ступеня зацікавленості щодо придбання комп'ютера