

## ІЄРАРХІЧНА КЛАСТЕРИЗАЦІЯ ДЛЯ ВИЗНАЧЕННЯ СУКУПНОСТІ ФУНКЦІОНАЛЬНИХ ТА НЕФУНКЦІОНАЛЬНИХ ВИМОГ КОМП'ЮТЕРНИХ СИСТЕМ

**Ievhenia Tysh PhD., Ruslan Shalapay**

### HIERARCHICAL CLUSTERIZATION FOR DETERMINING FUNCTIONAL AND NON-FUNCTIONAL REQUIREMENTS OF COMPUTER SYSTEMS

Ієрархічна кластеризація, як випливає з назви – це алгоритм, який формує ієрархію кластерів. Цей алгоритм починається з усіх елементів даних, які розглядаються як окремі кластери. Після цього два найближчі кластери об'єднуються в один. Зрештою, цей алгоритм завершується, коли залишається лише один кластер. Результати ієрархічної кластеризації можна показати за допомогою дендрограми. Дендрограму можна інтерпретувати як показано на рис. 1.

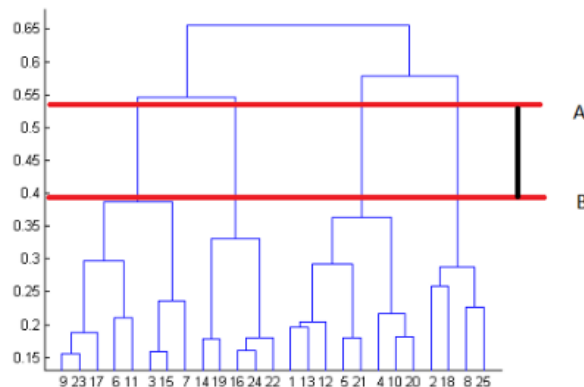


Рис. 1. Формування кластерів на основі відстані між ними

На рис. 1 показано вхідні елементи даних у вигляді 25 точок, кожна з яких представляє окремий кластер. Потім два найближчі кластери об'єднуються, доки не залишиться лише один кластер у верхній частині.

Висота на дендрограмі, на якій зливаються два кластери, представляє відстань між двома кластерами в просторі даних. Рішення щодо кількості кластерів, які найкраще відображають різні групи, можна вибрати, аналізуючи дендрограму. Найкращий вибір кількості кластерів визначається кількістю вертикальних ліній у дендрограмі, розрізаних горизонтальною лінією, яка може перетинати максимальну відстань по вертикалі, не перетинаючи кластер.

У наведеному на рис. 1 прикладі, найкращий вибір кількості кластерів буде 4, оскільки червона горизонтальна лінія на дендрограмі нижче покриває максимальну вертикальну відстань АВ.

Дві важливі речі, які потрібно знати про ієрархічну кластеризацію. Цей алгоритм реалізовано з використанням підходу знизу вгору. Також можна застосувати низхідний підхід, починаючи з того, що усі точки даних, задають один кластер.