

Варіант 1

Навести приклади застосування алгоритмів кластеризації. Загальна постановка задачі кластеризації.

Задача. Виконати кластеризацію наступних точок на два кластери, використовуючи алгоритм k-середніх, якщо початковим наближенням є точки (1,1) і (4,3), відповідно, та використано евклідову відстань:

№ точки	1	2	3	4	5	6	7	8	9
Координати	(1,1)	(1,2)	(2,1)	(2,2)	(2,3)	(3,3)	(4,2)	(4,3)	(5,3)

Значення відстані за формулами **здебільшого** рахувати не треба. Потрібно визначити до якого кластеру відстань мінімальна. У пропонованих задачах це **здебільшого** можна визначити графічно без розрахунку точних значень відстані.

Достатньо зробити декілька ітерацій методу k-середніх. Значення цільової функції рахувати не потрібно.

Варіант 2

Етапи агломеративного ієрархічного алгоритму найближчого сусіда.

Задача. Побудувати мінімальне покриваюче дерево для наступного графу, використовуючи алгоритм Крускала. Проілюструвати послідовність побудови дерева.

Вершина i	0	0	0	1	2	3	4	4	6	1
Вершина j	1	2	3	3	3	4	5	6	7	7
Вага ребра	6	17	11	25	8	2	21	14	9	19

Варіант 3

Як можна визначати відстань між кластерами в агломеративному ієрархічному алгоритмі?

Задача. Виконати кластеризацію наступних точок на два кластери, використовуючи алгоритм k-середніх, якщо початковим наближенням є точки (1,2) і (4,2), відповідно, та використано евклідову відстань:

№ точки	1	2	3	4	5	6	7	8	9
Координати	(1,1)	(1,2)	(2,1)	(2,2)	(2,3)	(3,3)	(4,2)	(4,3)	(5,2)

Значення відстані за формулами **здебільшого** рахувати не треба. Потрібно визначити до якого кластеру відстань мінімальна. У пропонованих задачах це **здебільшого** можна визначити графічно без розрахунку точних значень відстані.

Достатньо зробити декілька ітерацій методу k-середніх. Значення цільової функції рахувати не потрібно.

Варіант 4

Клас `sklearn.cluster.AgglomerativeClustering`.

Задача. Побудувати мінімальне покриваюче дерево для наступного графу, використовуючи алгоритм Прима. Проілюструвати послідовність побудови дерева.

Вершина i	0	0	0	1	2	3	4	4	6	1
Вершина j	1	2	3	3	3	4	5	6	7	7
Вага ребра	6	17	11	25	8	2	21	14	9	19

Варіант 5

Етапи розділяючого ієрархічного алгоритму DIANA.

Задача. Виконати кластеризацію наступних точок на два кластери, використовуючи алгоритм k-середніх, якщо початковим наближенням є точки (1,1) і (4,3), відповідно, та використано евклідову відстань:

№ точки	1	2	3	4	5	6	7	8	9
Координати	(1,3)	(1,2)	(2,4)	(2,2)	(2,3)	(3,3)	(4,2)	(4,3)	(5,3)

Значення відстані за формулами **здебільшого** рахувати не треба. Потрібно визначити до якого кластеру відстань мінімальна. У пропонованих задачах це **здебільшого** можна визначити графічно без розрахунку точних значень відстані.

Достатньо зробити декілька ітерацій методу k-середніх. Значення цільової функції рахувати не потрібно.

Варіант 6

Клас `sklearn.cluster.KMeans`.

Задача. Побудувати мінімальне покриваюче дерево для наступного графу, використовуючи алгоритм Крускала. Проілюструвати послідовність побудови дерева.

Вершина i	0	0	0	1	2	3	4	4	6	1
Вершина j	1	2	3	3	3	4	5	6	7	7
Вага ребра	26	10	9	15	18	20	21	14	9	19

Варіант 7

Етапи алгоритму нечітких k-середніх. Переваги і недоліки.

Задача. Побудувати мінімальне покриваюче дерево для наступного графу, використовуючи алгоритм Крускала. Проілюструвати послідовність побудови дерева.

Вершина i	0	0	0	1	2	3	4	4	6	1
Вершина j	1	2	3	3	3	4	5	6	7	7
Вага ребра	23	7	11	5	18	20	21	14	9	19

Варіант 8

Алгоритм G-середніх.

Задача. Виконати кластеризацію наступних точок на два кластери, використовуючи алгоритм k-середніх, якщо початковим наближенням є точки (1,2) і (4,4), відповідно, та використано евклідову відстань:

№ точки	1	2	3	4	5	6	7	8	9
Координати	(1,4)	(1,2)	(2,4)	(2,2)	(2,3)	(3,3)	(4,4)	(4,3)	(5,3)

Значення відстані за формулами **здебільшого** рахувати не треба. Потрібно визначити до якого кластеру відстань мінімальна. У пропонованих задачах це **здебільшого** можна визначити графічно без розрахунку точних значень відстані.

Достатньо зробити декілька ітерацій методу k-середніх. Значення цільової функції рахувати не потрібно.

Варіант 9

Навести приклади застосування алгоритмів кластеризації. Постановка задачі розділу суміші.

Задача. Виконати кластеризацію наступних точок на два кластери, використовуючи алгоритм k-середніх, якщо початковим наближенням є точки (0,0) і (5,3), відповідно, та використано евклідову відстань:

№ точки	1	2	3	4	5	6	7	8	9
Координати	(1,1)	(1,2)	(2,1)	(2,2)	(2,0)	(3,3)	(4,2)	(4,3)	(5,4)

Значення відстані за формулами **здебільшого** рахувати не треба. Потрібно визначити до якого кластеру відстань мінімальна. У пропонованих задачах це **здебільшого** можна визначити графічно без розрахунку точних значень відстані.

Достатньо зробити декілька ітерацій методу k-середніх. Значення цільової функції рахувати не потрібно.

Варіант 10

Детально описати E-крок базового алгоритму Expectation-Maximization

Задача. Побудувати мінімальне покриваюче дерево для наступного графу, використовуючи алгоритм Крускала. Проілюструвати послідовність побудови дерева.

Вершина i	0	0	0	1	2	3	4	4	6	1
Вершина j	1	2	3	3	3	4	5	6	7	7
Вага ребра	26	10	9	15	18	20	21	14	9	19

Варіант 11

Детально описати М-крок базового алгоритму Expectation-Maximization

Задача. Виконати кластеризацію наступних точок на два кластери, використовуючи алгоритм k-середніх, якщо початковим наближенням є точки (1,4) і (4,1), відповідно, та використано евклідову відстань:

№ точки	1	2	3	4	5	6	7	8	9
Координати	(1,1)	(1,2)	(1,3)	(3,2)	(3,3)	(3,4)	(4,2)	(4,3)	(4,4)

Значення відстані за формулами **здебільшого** рахувати не треба. Потрібно визначити до якого кластеру відстань мінімальна. У пропонованих задачах це **здебільшого** можна визначити графічно без розрахунку точних значень відстані.

Достатньо зробити декілька ітерацій методу k-середніх. Значення цільової функції рахувати не потрібно.

Варіант 12

Етапи алгоритму Expectation-Maximization з фіксованою кількістю компонент.

Задача. Побудувати мінімальне покриваюче дерево для наступного графу, використовуючи алгоритм Крускала. Проілюструвати послідовність побудови дерева.

Вершина i	0	0	0	1	2	3	4	4	6	1
Вершина j	1	2	3	3	3	4	5	6	7	7
Вага ребра	23	7	11	5	18	20	21	14	9	19

Варіант 13

Стохастичний алгоритм Expectation-Maximization: ідея та опис трьох етапів.

Задача. Виконати кластеризацію наступних точок на два кластери, використовуючи алгоритм k-середніх, якщо початковим наближенням є точки (2,0) і (4,4), відповідно, та використано евклідову відстань:

№ точки	1	2	3	4	5	6	7	8	9
Координати	(1,1)	(1,2)	(2,1)	(2,2)	(3,1)	(3,2)	(4,4)	(5,3)	(5,4)

Пояснення. Значення відстані за формулами **здебільшого** рахувати не треба. Потрібно визначити до якого кластеру відстань мінімальна. У пропонованих задачах це **здебільшого** можна визначити графічно без розрахунку точних значень відстані.

Достатньо зробити декілька ітерацій методу k-середніх. Значення цільової функції рахувати не потрібно.

Варіант 14

Описати що дано і що потрібно визначити в алгоритмі Expectation-Maximization з послідовним додаванням компонент. Яке призначення параметрів в цьому алгоритмі?

Задача. Побудувати мінімальне покриваюче дерево для наступного графу, використовуючи алгоритм Прима. Проілюструвати послідовність побудови дерева.

Вершина i	0	0	0	1	2	3	4	4	6	1
Вершина j	1	2	3	3	3	4	5	6	7	7
Вага ребра	26	10	9	15	18	20	21	14	9	19

Варіант 15

Навести приклади застосування алгоритмів кластеризації. Описати алгоритм знаходження зв'язних компонент, його переваги і недоліки.

Задача. Виконати кластеризацію наступних точок на три кластери, використовуючи алгоритм k-середніх, якщо початковим наближенням є точки (0,0), (3,3) і (6,1), відповідно, та використано евклідову відстань:

№ точки	1	2	3	4	5	6	7	8	9
Координати	(1,1)	(1,2)	(2,1)	(2,2)	(3,4)	(3,5)	(5,1)	(5,2)	(4,2)

Пояснення. Значення відстані за формулами **здебільшого** рахувати не треба. Потрібно визначити до якого кластеру відстань мінімальна. У пропонованих задачах це **здебільшого** можна визначити графічно без розрахунку точних значень відстані.

Достатньо зробити декілька ітерацій методу k-середніх. Значення цільової функції рахувати не потрібно.

Варіант 16

Поняття мінімального покриваючого дерева, навести приклад. Що таке переріз графу та перехресне ребро перерізу ?

Задача. Виконати кластеризацію наступних точок на два кластери, використовуючи алгоритм k-середніх, якщо початковим наближенням є точки (1,1) і (4,3), відповідно, та використано евклідову відстань:

№ точки	1	2	3	4	5	6	7	8	9
Координати	(1,1)	(1,2)	(2,1)	(2,2)	(2,3)	(3,3)	(4,2)	(4,3)	(5,3)

Пояснення. Значення відстані за формулами **здебільшого** рахувати не треба. Потрібно визначити до якого кластеру відстань мінімальна. У пропонованих задачах це **здебільшого** можна визначити графічно без розрахунку точних значень відстані.

Достатньо зробити декілька ітерацій методу k-середніх. Значення цільової функції рахувати не потрібно.

Варіант 17

«Жадібний» алгоритм побудови мінімального покриваючого дерева. Які переваги і недоліки цього алгоритму?

Задача. Виконати кластеризацію наступних точок на три кластери, використовуючи алгоритм k-середніх, якщо початковим наближенням є точки (0,0), (3,2) і (6,1), відповідно, та використано евклідову відстань:

№ точки	1	2	3	4	5	6	7	8	9
Координати	(1,1)	(1,3)	(2,1)	(2,2)	(3,4)	(3,5)	(6,0)	(6,1)	(5,2)

Пояснення. Значення відстані за формулами **здебільшого** рахувати не треба. Потрібно визначити до якого кластеру відстань мінімальна. У пропонованих задачах це **здебільшого** можна визначити графічно без розрахунку точних значень відстані.

Достатньо зробити декілька ітерацій методу k-середніх. Значення цільової функції рахувати не потрібно.

Варіант 18

Клас `sklearn.cluster.KMeans`.

Задача. Побудувати мінімальне покриваюче дерево для наступного графу, використовуючи алгоритм Крускала. Проілюструвати послідовність побудови дерева.

Вершина i	0	0	0	1	2	3	4	4	6	1
Вершина j	1	2	3	3	3	4	5	6	7	7
Вага ребра	26	10	9	15	18	20	21	14	9	19

Варіант 19

Етапи алгоритму нечітких k-середніх. Переваги і недоліки.

Задача. Побудувати мінімальне покриваюче дерево для наступного графу, використовуючи алгоритм Крускала. Проілюструвати послідовність побудови дерева.

Вершина i	0	0	0	1	2	3	4	4	6	1
Вершина j	1	2	3	3	3	4	5	6	7	7
Вага ребра	23	7	11	5	18	20	21	14	9	19

Варіант 20

Описати метод середнього зсуву. Що таке ядро в цьому методі?

Задача. Виконати кластеризацію наступних точок на три кластери, використовуючи алгоритм k-середніх, якщо початковим наближенням є точки (2,2), (4,5) і (5,3), відповідно, та використано евклідову відстань:

№ точки	1	2	3	4	5	6	7	8	9	10
Координати	(1,2)	(1,3)	(2,3)	(2,4)	(4,2)	(4,3)	(4,4)	(5,5)	(5,6)	(6,5)

Значення відстані за формулами **здебільшого** рахувати не треба. Потрібно визначити до якого кластеру відстань мінімальна. У пропонованих задачах це **здебільшого** можна визначити графічно без розрахунку точних значень відстані.

Достатньо зробити декілька ітерацій методу k-середніх. Значення цільової функції рахувати не потрібно.

Варіант 21

Описати алгоритм DBSCAN.

Задача. Виконати кластеризацію наступних точок на три кластери, використовуючи алгоритм k-середніх, якщо початковим наближенням є точки (2,2), (4,5) і (5,3), відповідно, та використано евклідову відстань:

№ точки	1	2	3	4	5	6	7	8	9	10
Координати	(1,1)	(1,3)	(2,3)	(2,4)	(3,3)	(4,3)	(4,4)	(5,5)	(5,6)	(6,5)

Значення відстані за формулами **здебільшого** рахувати не треба. Потрібно визначити до якого кластеру відстань мінімальна. У пропонованих задачах це **здебільшого** можна визначити графічно без розрахунку точних значень відстані.

Достатньо зробити декілька ітерацій методу k-середніх. Значення цільової функції рахувати не потрібно.

Варіант 22

Підбір значень параметрів в DBSCAN. Що таке крива K-distance plot і для чого вона використовується? Навести приклад.

Задача. Побудувати мінімальне покриваюче дерево для наступного графу, використовуючи алгоритм Прима. Проілюструвати послідовність побудови дерева.

Вершина i	0	0	0	1	2	3	4	4	6	1
Вершина j	1	2	3	3	3	4	5	6	7	7
Вага ребра	23	7	11	5	18	20	21	14	9	19

Варіант 23

Навести приклади застосування алгоритмів кластеризації. Ідея базового алгоритму FOREL.

Задача. Побудувати мінімальне покриваюче дерево для наступного графу, використовуючи алгоритм Крускала. Проілюструвати послідовність побудови дерева.

Вершина i	0	0	0	1	2	3	4	4	6	1
Вершина j	1	2	3	3	3	4	5	6	7	7
Вага ребра	6	17	11	25	8	2	21	14	9	19

Варіант 24

Етапи базового алгоритму FOREL.

Задача. Виконати кластеризацію наступних точок на три кластери, використовуючи алгоритм k-середніх, якщо початковим наближенням є точки (2,3), (4,5) і (5,3), відповідно, та використано евклідову відстань:

№ точки	1	2	3	4	5	6	7	8	9	10
Координати	(1,2)	(1,3)	(2,3)	(2,1)	(4,2)	(4,3)	(3,3)	(5,5)	(5,6)	(6,5)

Пояснення. Значення відстані за формулами **здебільшого** рахувати не треба. Потрібно визначити до якого кластеру відстань мінімальна. У пропонованих задачах це **здебільшого** можна визначити графічно без розрахунку точних значень відстані.

Достатньо зробити декілька ітерацій методу k-середніх. Значення цільової функції рахувати не потрібно.

Варіант 25

Описати переваги і недоліки базового алгоритму FOREL.

Задача. Побудувати мінімальне покриваюче дерево для наступного графу, використовуючи алгоритм Прима. Проілюструвати послідовність побудови дерева.

Вершина i	0	0	0	1	2	3	4	4	6	1
Вершина j	1	2	3	3	3	4	5	6	7	7
Вага ребра	23	7	11	5	18	20	21	14	9	19

Варіант 26

Метрики якості кластеризації: Homogeneity, Completeness, V-measure.

Задача. Побудувати мінімальне покриваюче дерево для наступного графу, використовуючи алгоритм Прима. Проілюструвати послідовність побудови дерева.

Вершина i	0	0	0	1	2	3	4	4	6	1
Вершина j	1	2	3	3	3	4	5	6	7	7
Вага ребра	6	17	11	25	8	2	21	14	9	19

Варіант 27

Метрика якості кластеризації - коефіцієнт силуету Silhouette Coefficient.

Задача. Виконати кластеризацію наступних точок на два кластери, використовуючи алгоритм k-середніх, якщо початковим наближенням є точки (1,4) і (4,1), відповідно, та використано евклідову відстань:

№ точки	1	2	3	4	5	6	7	8	9
Координати	(1,1)	(1,2)	(1,3)	(3,2)	(3,3)	(3,4)	(4,2)	(4,3)	(4,4)

Пояснення. Значення відстані за формулами **здебільшого** рахувати не треба. Потрібно визначити до якого кластеру відстань мінімальна. У пропонованих задачах це **здебільшого** можна визначити графічно без розрахунку точних значень відстані.

Достатньо зробити декілька ітерацій методу k-середніх. Значення цільової функції рахувати не потрібно.