

Лабораторна робота №7

ДОСЛІДЖЕННЯ МЕТОДІВ НЕКОНТРОЛЬОВАНОГО НАВЧАННЯ

Мета: використовуючи спеціалізовані бібліотеки та мову програмування Python дослідити методи неконтрольованої класифікації даних у машинному навчанні.

Хід роботи

Завдання 2.1. Кластеризація даних за допомогою методу k-середніх

Провести кластеризацію даних методом k-середніх. Використовувати файл вхідних даних: data_clustering.txt.

Лістинг коду:

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
from sklearn import metrics

# Завантажимо вхідні дані з файлу.
X = np.loadtxt('data_clustering.txt', delimiter=',')

# Щоб застосувати k-середніх необхідно задати кількість кластерів
num_clusters = 5

# Візуалізуємо вхідні дані, щоб побачити, як виглядає розподіл
plt.scatter(X[:, 0], X[:, 1])
plt.title('Вхідні дані')
plt.xlabel('Feature 1')
plt.ylabel('Feature 2')
plt.show()
plt.figure()
plt.scatter(X[:, 0], X[:, 1], marker='o', facecolors='none', edgecolors='black',
s=80)

x_min, x_max = X[:, 0].min() - 1, X[:, 0].max() + 1
y_min, y_max = X[:, 1].min() - 1, X[:, 1].max() + 1

plt.title('Вхідні дані')
plt.xlim(x_min, x_max)
plt.ylim(y_min, y_max)
plt.xticks(())
plt.yticks(())
# Створення об'єкту KMeans
kmeans = KMeans(init='k-means++', n_clusters=num_clusters, n_init=10)

# Навчимо модель k-середніх на вхідних даних.
kmeans.fit(X)

# Щоб візуалізувати межі, ми маємо створити сітку точок та обчислити модель на
всіх вузлах сітки. Визначимо крок сітки.
```

					ДУ «Житомирська політехніка».24.121.16.000 – Лр7			
Змн.	Арк.	№ докум.	Підпис	Дата	Звіт з лабораторної роботи	Лім.	Арк.	Аркушів
Розроб.		Некритий В.Ю.						
Перевір.		Іванов Д.А.					1	4
Керівник						ФІКТ Гр. ІПЗ-21-5		
Н. контр.								
Зав. каф.								

```

step_size = 0.01

# Далі визначимо саму сітку і переконаємось в тому, що вона охоплює всі вхідні
значення.
# Відображення точок сітки
x_min, x_max = X[:, 0].min() - 1, X[:, 0].max() + 1
y_min, y_max = X[:, 1].min() - 1, X[:, 1].max() + 1

x_vals, y_vals = np.meshgrid(np.arange(x_min, x_max, step_size), np.arange(y_min,
y_max, step_size))

# Спроекуємо результати всіх точок сітки, використовуючи навчену модель k-
середніх.
# Передбачення вихідних міток для всіх точок сітки
output = kmeans.predict(np.c_[x_vals.ravel(), y_vals.ravel()])

# Відобразить на графіку вихідні значення та виділить кожен область своїм
кольором.
# Графічне відображення областей та виділення їх кольором
output = output.reshape(x_vals.shape)
plt.figure()
plt.clf()
plt.imshow(output, interpolation='nearest',
            extent=(x_vals.min(), x_vals.max(),
                    y_vals.min(), y_vals.max()),
            cmap=plt.cm.Paired, aspect='auto',
            origin='lower')

# Відобразить вихідні дані на виділених кольором областях.
# Відображення вихідних точок
plt.scatter(X[:, 0], X[:, 1], marker='o', facecolors='none',
            edgecolors='black', s=80)

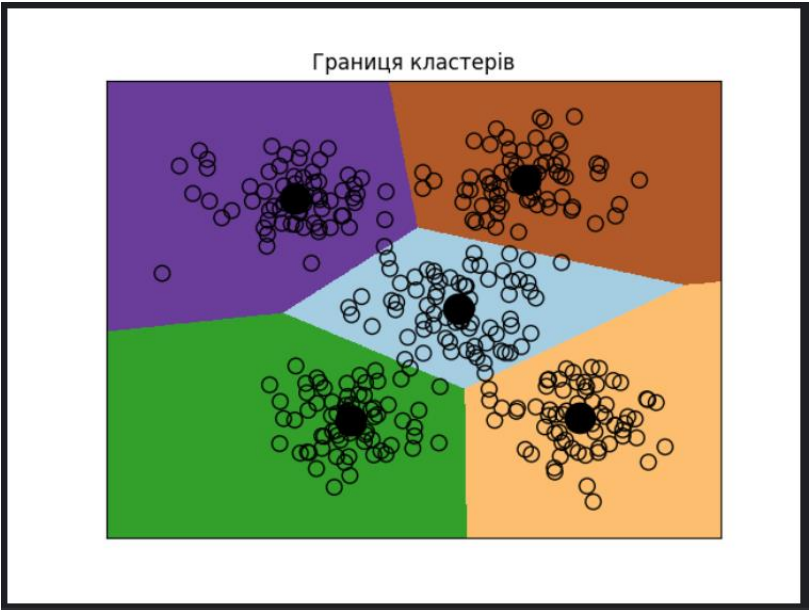
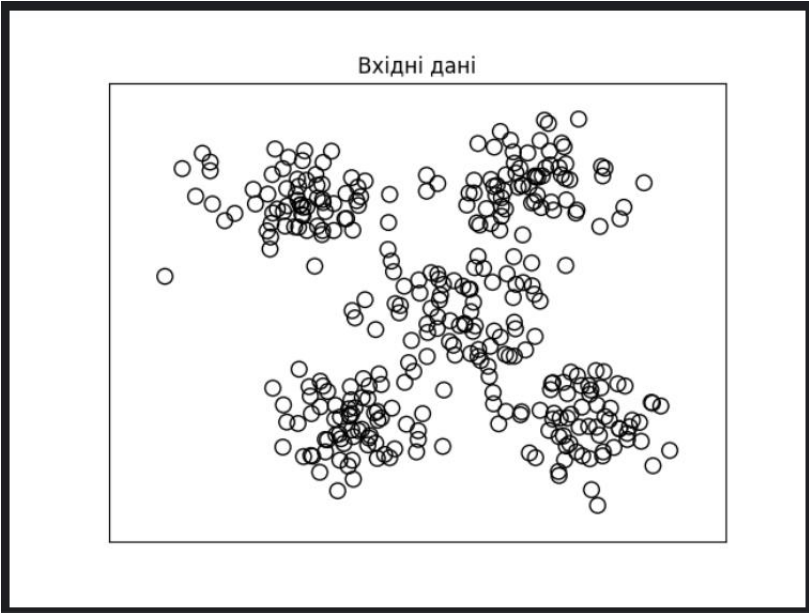
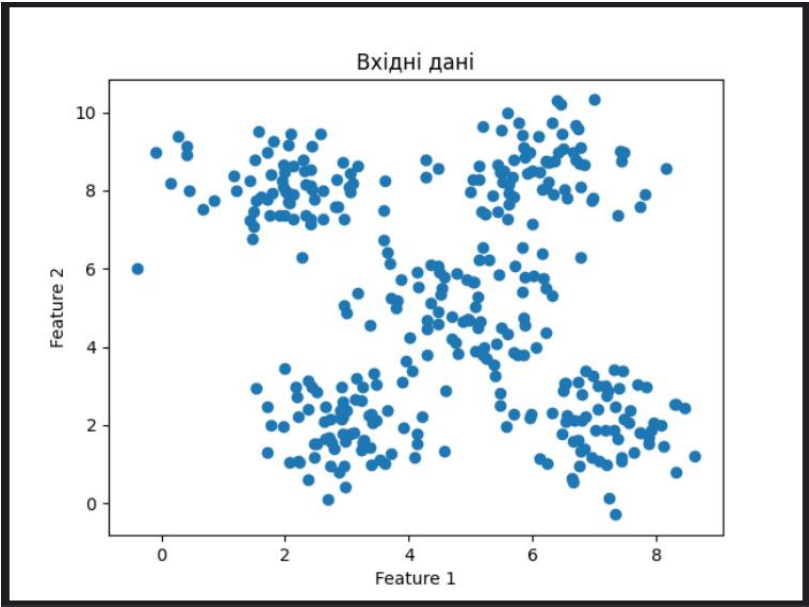
# Відобразить на графіку центри кластерів, отримані з використанням методу k-
середніх.
# Відображення центрів кластерів
cluster_centers = kmeans.cluster_centers_
plt.scatter(cluster_centers[:, 0], cluster_centers[:, 1],
            marker='o', s=210, linewidths=4, color='black',
            zorder=12, facecolors='black')

x_min, x_max = X[:, 0].min() - 1, X[:, 0].max() + 1
y_min, y_max = X[:, 1].min() - 1, X[:, 1].max() + 1
plt.title('Границя кластерів')
plt.xlim(x_min, x_max)
plt.ylim(y_min, y_max)
plt.xticks(())
plt.yticks(())
plt.show()

```

Результат виконання:

		Некритий В.Ю.			ДУ «Житомирська політехніка».24.121.16.000 – Лр7	Арк.
		Іванов Д.А.				
Змн.	Арк.	№ докум.	Підпис	Дата		2



Висновок: використовуючи спеціалізовані бібліотеки та мову програмування Python дослідив методи неконтрольованої класифікації даних у машинному навчанні.

		Некритий В.Ю.			ДУ «Житомирська політехніка».24.121.16.000 – Лр7	Арк.
		Іванов Д.А.				
Змн.	Арк.	№ докум.	Підпис	Дата		4