

МІНІСТЕРСТВО ОСВІТИ ТА НАУКИ УКРАЇНИ
Національний технічний університет
«Харківський політехнічний інститут»
Кафедра ГМКГ

Лабораторна работа №5

З дисципліни «Інтелектуальний аналіз даних»

Виконав:

Студент групи ІКМ-220 г.

Ульянов Кирило Юрійович

Перевірив:

Доц. Дашкевич А.О.

Харків 2023

Мета роботи: : вивчення базових алгоритмів зниження розмірності для задач кластеризації та візуалізації даних.

Завдання на роботу: завантаження набору даних, зниження розмірності даних лінійними та нелінійними методами, знаходження способу зниження розмірності та оптимальної розмірності для розв'язання задачі кластеризації на наборі із меншою розмірністю, порівняльний аналіз лінійного та нелінійних алгоритмів.

Варіант: 20

$N = 100$

Метод: Isomap

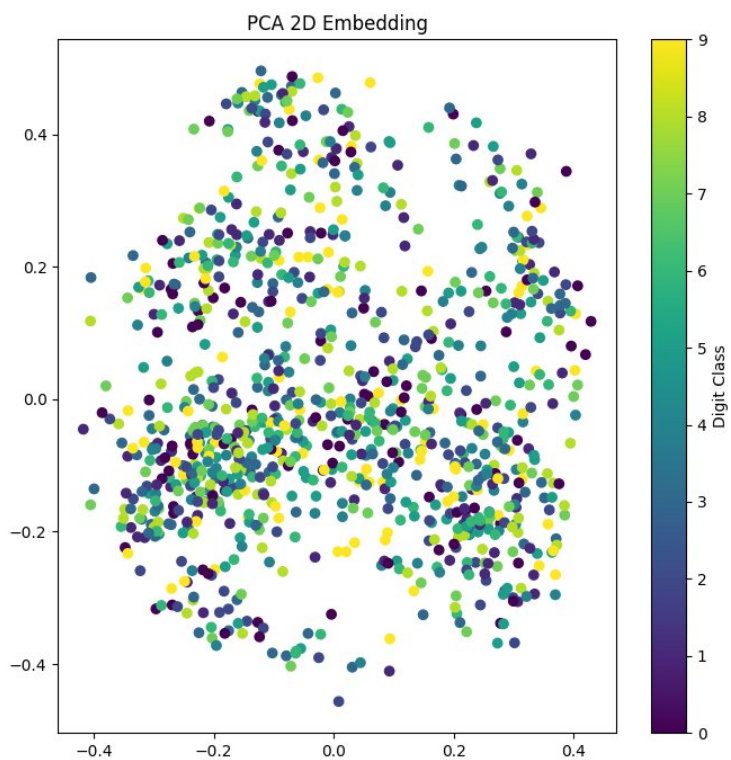
$D1 = 5$

$D2 = 15$

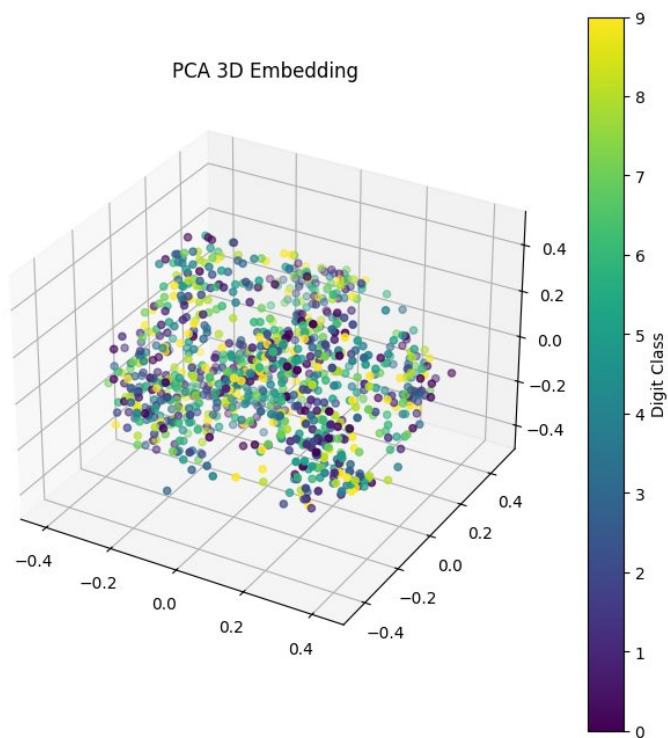
1) Візуалізація даних з якими будемо працювати.



2) Вкладення даних у простори розмірності 2 та 3 за допомогою методу головних компонентів (PCA)

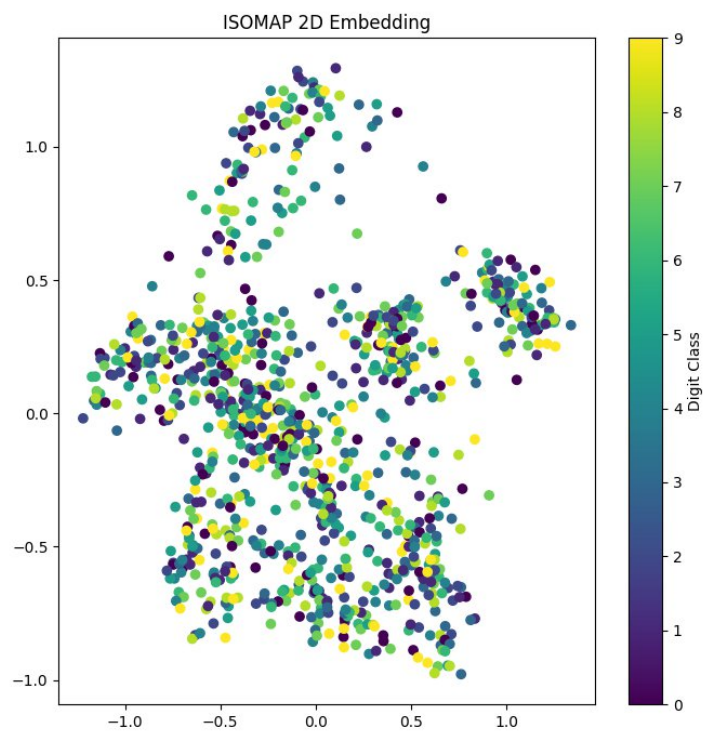


Візуалізація у 2-вимірному просторі

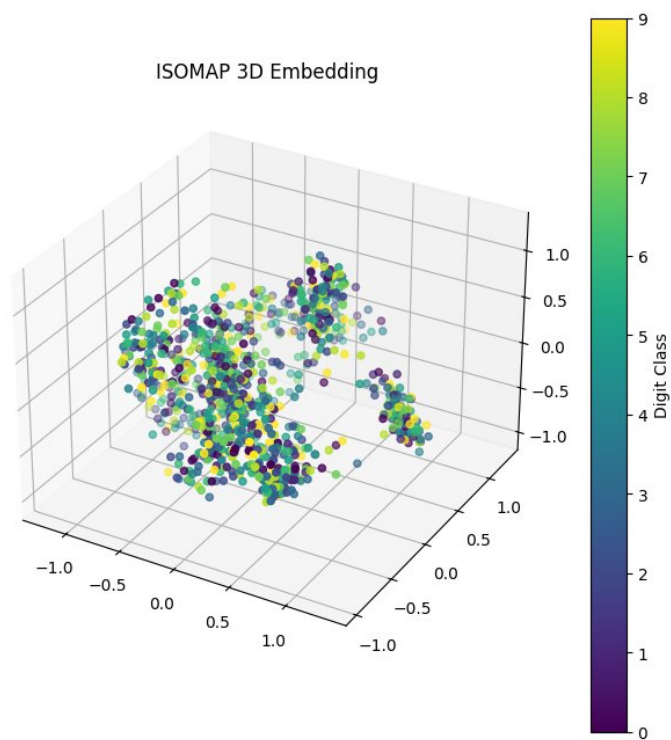


Візуалізація у 3-вимірному просторі

3) Вкладення даних у простори розмірності 2 та 3 за допомогою методу (ISOMAP)

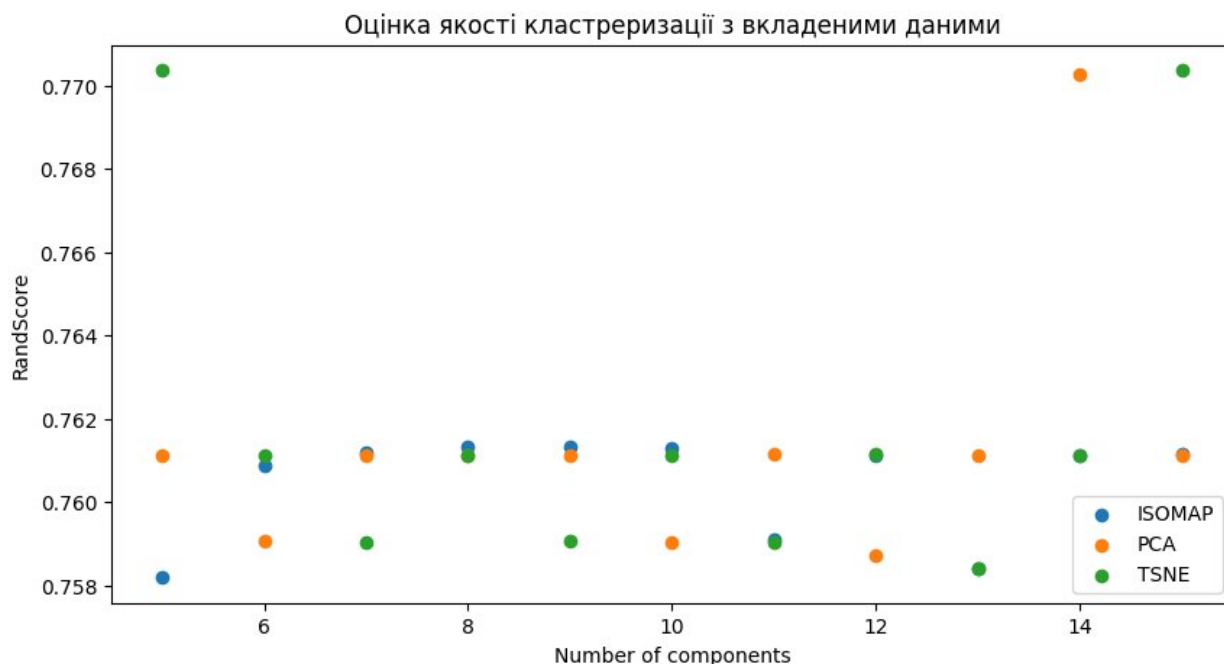


Візуалізація у 2-вимірному просторі



Візуалізація у 3-вимірному просторі

4) Порівняльна таблиця результатів кластеризації вкладених даних за допомогою різних методів зниження розмірності даних.



Кількість компонентів	точність з PCA	точність з ISOMAP	точність з T-SNE
5	0.7611	0.7582	0.7704
6	0.7591	0.7609	0.7611
7	0.7611	0.7612	0.7590
8	0.7611	0.7613	0.7611
9	0.7611	0.7613	0.7591
10	0.7590	0.7613	0.7611
11	0.7612	0.7591	0.7590
12	0.7587	0.7611	0.7612
13	0.7611	0.7584	0.7584
14	0.7703	0.7611	0.7611
15	0.7611	0.7612	0.7704

Можна побачити що в цілому не сильно змінювалася точність алгоритмів на визначеному діапазоні. Але все ж таки на деяких проміжках найкращу точність отримали алгоритми TSNE та PCA. Помітно, що при зміні кількості компонент не сильно змінюється точність кластеризатору KMeans.