

Задание 3-4.

1. Выполнить визуализацию в двухмерном пространстве набора данных, указанного в таблице 1, используя алгоритмы нелинейного снижения размерности t-sne, UMAP, TriMAP и PacMAP.

Для алгоритмов t-sne, UMAP, TriMAP и PacMAP имеются программные реализации в Python:

t-sne (<https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html>),

UMAP (<https://umap-learn.readthedocs.io/en/latest/>),

TriMap (<https://pypi.org/project/trimap/>)

и PacMAP (<https://pypi.org/project/pacmap/>).

Для каждого алгоритма рассмотреть не менее 6 вариантов сочетаний значений их параметров.

Кроме того, рассмотреть результаты вложения набора данных в двухмерное пространство при значениях параметров алгоритмов, заданных по умолчанию.

При работе с набором данных реализовать различные варианты масштабирования:

MinMax

(<https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html?highlight=minmax#sklearn.preprocessing.MinMaxScaler>),

Standard

(<https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html#sklearn.preprocessing.StandardScaler>),

Robust

(<https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html?highlight=robust#sklearn.preprocessing.RobustScaler>).

Выполнить сравнительный анализ результатов масштабирования (https://scikit-learn.org/stable/auto_examples/preprocessing/plot_all_scaling.html#sphx-glr-auto-examples-preprocessing-plot-all-scaling-py, <https://www.geeksforgeeks.org/standardscaler-minmaxscaler-and-robustscaler-techniques-ml/>).

Разбиение на обучающую и тестовую выборку не проводить.

Библиотека алгоритмов машинного обучения:

<https://scikit-learn.org/stable/>.

Таблица 1. Варианты заданий

Вариант	Источник набора данных
1	http://archive.ics.uci.edu/ml/datasets/Post+Operative+Patient
2	http://archive.ics.uci.edu/ml/datasets/Primary+Tumor
3	http://archive.ics.uci.edu/ml/datasets/Soybean+%28Large%29
4	http://archive.ics.uci.edu/ml/datasets/Low+Resolution+Spectrometer
5	http://archive.ics.uci.edu/ml/datasets/Ionosphere
6	http://archive.ics.uci.edu/ml/datasets/Horse+Colic
7	http://archive.ics.uci.edu/ml/datasets/Hepatitis
8	http://archive.ics.uci.edu/ml/datasets/Heart+Disease
9	http://archive.ics.uci.edu/ml/datasets/Hayes-Roth
10	http://archive.ics.uci.edu/ml/datasets/Haberman%27s+Survival
11	http://archive.ics.uci.edu/ml/datasets/Glass+Identification
12	http://archive.ics.uci.edu/ml/datasets/Flags
13	http://archive.ics.uci.edu/ml/datasets/Ecoli
14	http://archive.ics.uci.edu/ml/datasets/Echocardiogram
15	http://archive.ics.uci.edu/ml/datasets/Dermatology
16	http://archive.ics.uci.edu/ml/datasets/Credit+Approval
17	http://archive.ics.uci.edu/ml/datasets/Pittsburgh+Bridges
18	http://archive.ics.uci.edu/ml/datasets/Spambase
19	http://archive.ics.uci.edu/ml/datasets/University
20	http://archive.ics.uci.edu/ml/datasets/statlog+(australian+credit+approval)

2. Реализовать вложение трехмерного объекта в двухмерное пространство.

Применить к набору данных mammoth.csv, описывающего трехмерный объект, при разных сочетаниях значений их параметров. Для каждого алгоритма рассмотреть не менее 6 вариантов сочетаний значений их параметров. Использовать различные варианты масштабирования данных.

Кроме того, рассмотреть результаты вложения трехмерного объекта в двухмерное пространство при значениях параметров алгоритмов, заданных по умолчанию.