

## Работа №4. Алгоритмы визуализации и снижения размерности

### Задачи:

1. Выбрать набор данных с размеченными объектами, который ранее использовался в той или иной работе курса.  
Построить диаграмму распределения объектов по классам. В случае необходимости провести предобработку.

2. Алгоритм самоорганизующихся карт Кохонена (рекомендуется использовать пакет minisom).

Построить модель Кохонена и разные варианты визуализации карты:

- распределение объектов по ячейкам карты;
- распределение объектов по классам для каждой ячейки

Сравнить объекты двух ячеек карты по средним значениям показателей.

3. Алгоритм t-SNE.

Выполнить преобразование данных в двумерное пространство с помощью алгоритма t-SNE и отобразить распределение объектов (цвет соответствует классу объекта).

4. Задание по варианту.

Вариант определяем по формуле  $V = 1 + (\text{ord}(c1) + \text{ord}(c2)) \% 4$

Вариант	Задание
1	Анализ расстояний между объектами одного и разных классов Выберите несколько пар объектов, относящихся к одному классу, и несколько пар объектов, относящихся к разным классам. Вычислите расстояния между объектами в исходном пространстве и двумерном пространстве (после применения алгоритма t-SNE). Определите ячейки карты Кохонена, соответствующие объектам.
2	Сопоставление результатов кластеризации kMeans и SOM Проведите кластеризацию алгоритмом kMeans, где параметр k равен числу ячеек карты SOM. Вычислите оценку согласованности кластерных решений Rand для kMeans и SOM. Проведите эксперименты с разными установками для SOM (параметры карты, инициализация).
3	Сопоставление результатов кластеризации с применением t-SNE Проведите кластеризацию с помощью алгоритма kMeans исходных (многомерных) данных и преобразованных данных (после алгоритма t-SNE). Вычислите оценку согласованности кластерных решений Rand. Проведите эксперименты с разными установками для t-SNE.
4	Анализ расстояний между объектами одной и разных ячеек карты Выберите 2 несмежные непустые ячейки карты Кохонена. Вычислите стат. характеристики (мин., средн., макс.) по расстояниям в двумерном пространстве t-SNE для пар объектов, относящихся к одной ячейке, и для пар объектов, относящихся к разным ячейкам.