МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ**

**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ «МИСиС»**

Институт ИТАСУ

Группа: **МПИ-20-4-2**

**ОТЧЕТ**

**по лабораторной работе №3**

**по курсу «Нейронные сети»**

**Выполнил: Хабибулин М.И.  
группа МПИ-20-4-2**

**Проверил: Курочкин И.И.**

**Москва 2020**

**Инструментарий:**

Язык программирования python 3.7

Библиотеки: matplotlib, numpy, random, pandas.

**Реализация:**

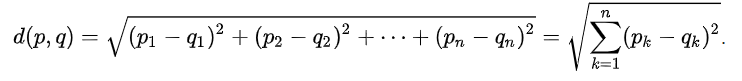
В ходе работы был реализован метод кластерного анализа K-means, а также алгоритм обучения самоорганизующейся карты Кохонена (SOM).

Реализована нейронная сеть с обучением без учителя. Карта SOM создается с помощью нескольких этапов:

1. Инициализация начальных весов случайным образом;
2. Обучение без учителя. Осуществляется путем последовательной коррекции векторов (векторами являются нейроны). На каждом шаге обучения из исходного набора данным случайно выбирается один из векторов, а затем производится поиск наиболее похожего на него вектора коэффициентов нейронов. Затем выбирается нейрон-победитель, расстояния до которого наименьшее. После этого корректируются веса нейрона-победителя и рядом стоящих нейронов в заданном радиусе.

Расстояние подсчитывалось двумя методами:

1. Евклидово расстояние рассчитывается по формуле



Где *p, q* – некоторые точки.

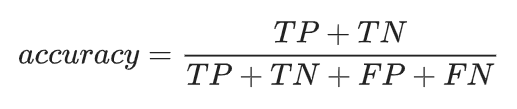
1. Расстояние Хэмминга рассчитывается по формуле

*.*

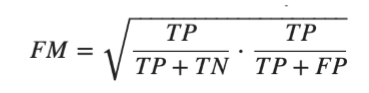
Метрики оценки качества кластеризации:

|  |  |
| --- | --- |
| True positive (TP) | False positive (FP) |
| False negative (FN) | True negative (TN) |

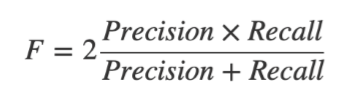
1. Accuracy рассчитывается по формуле



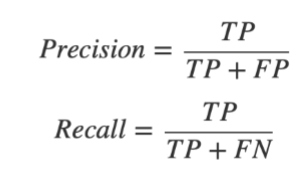
1. Folkes and Mallows Index рассчитывается по формуле



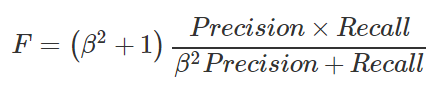
1. F-мера сбалансированная рассчитывается по формуле



Где



1. F-мера с приоритетом точности рассчитывается по формуле



Где .

1. F-мера с приоритетом полноты, где

**Результат работы:**

**Пример 1.**

3 линейно разделимых класса, находящихся далеко друг от друга

Параметры:

k=3

som\_size=10

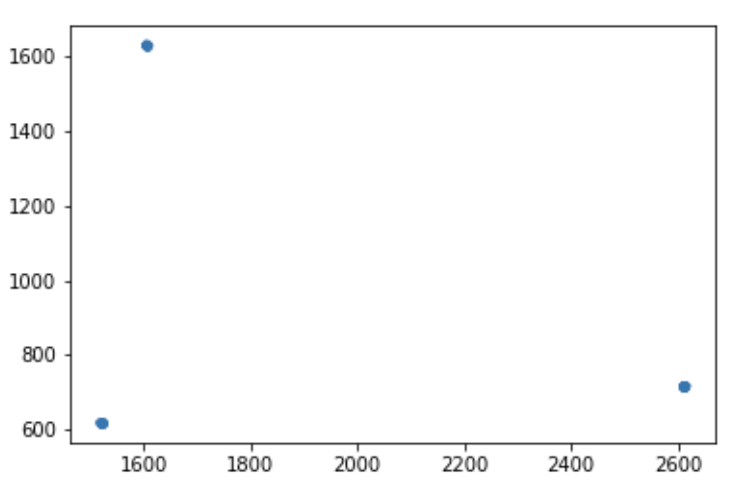
dimension = 2

learning\_rate=0.05

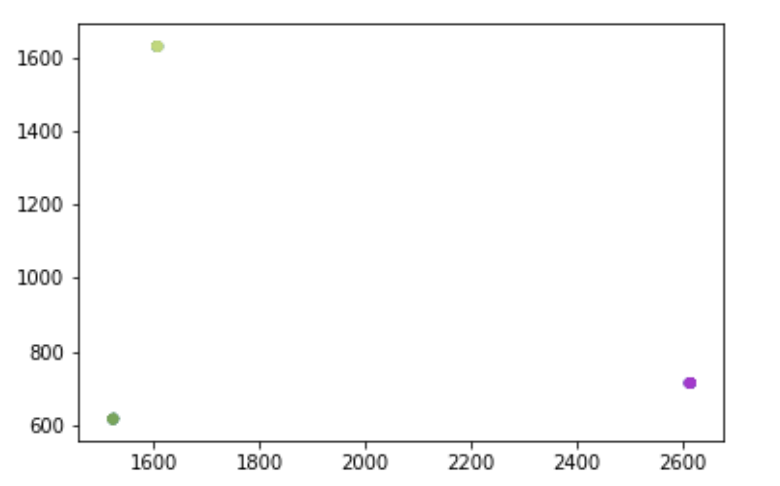
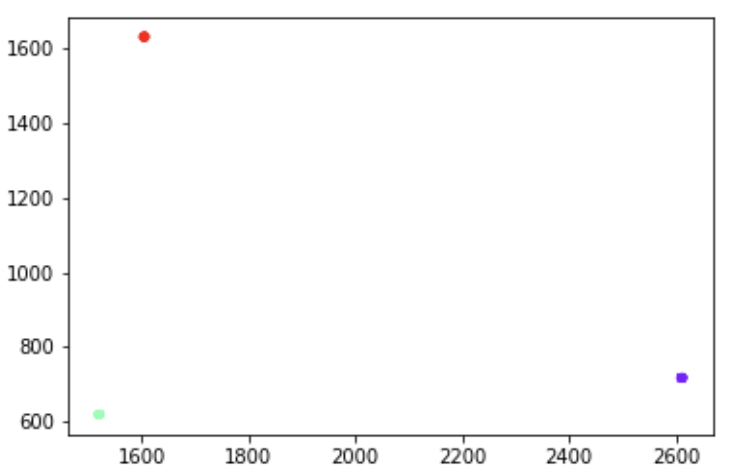
cooperation=5

epoch = 2000

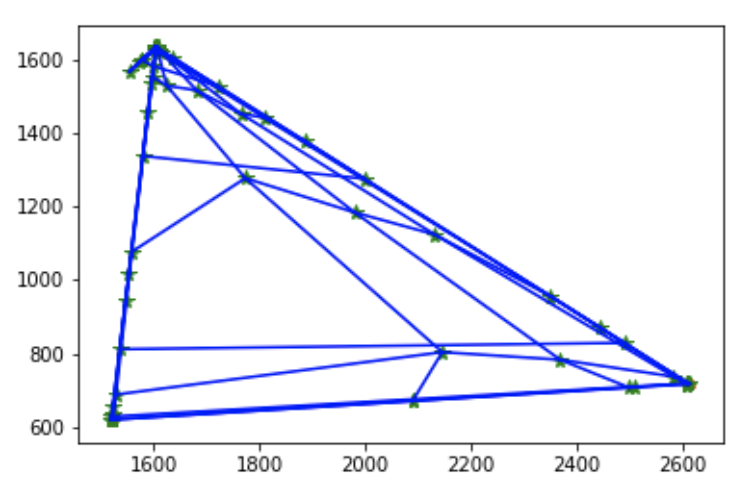
До кластеризации



После кластеризации

Слева - кластеризация K-means, справа - кластеризация SOM



SOM карта

**Сравнительная таблица 1.**

|  |  |  |
| --- | --- | --- |
| Метрика | K-means | Метод SOM |
| Rand Index | 1.0 | 1.0 |
| F1-measure | 1.0 | 1.0 |
| Fowlkes-Mallows Index | 1.0 | 1.0 |

**Пример 2.**

30 линейно разделимых класса, находящихся далеко друг от друга

Параметры:

k=30

som\_size=20

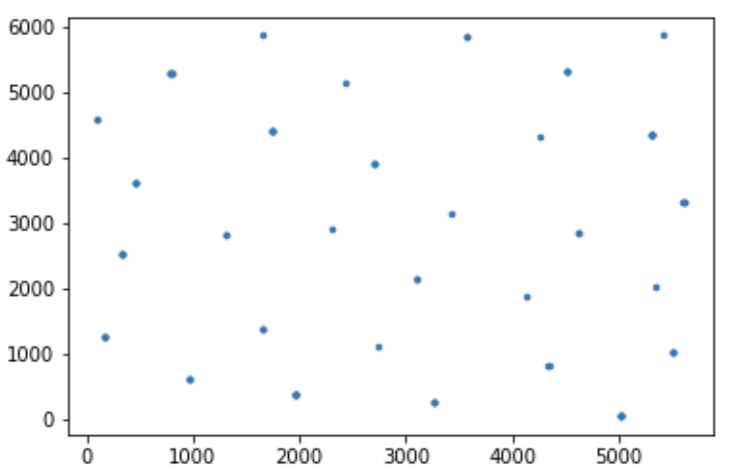
dimension = 2

learning\_rate=0.05

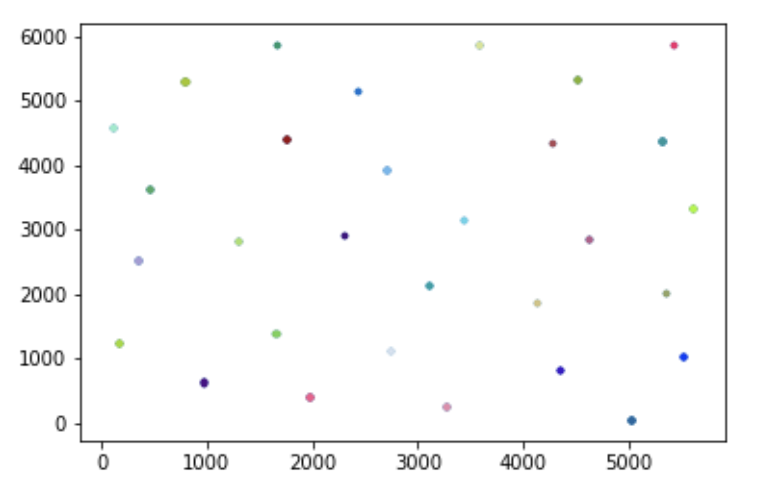
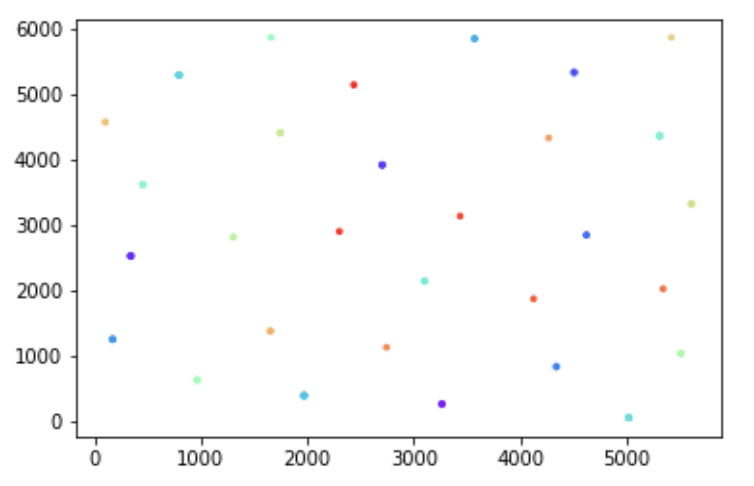
cooperation=6

epoch = 20000

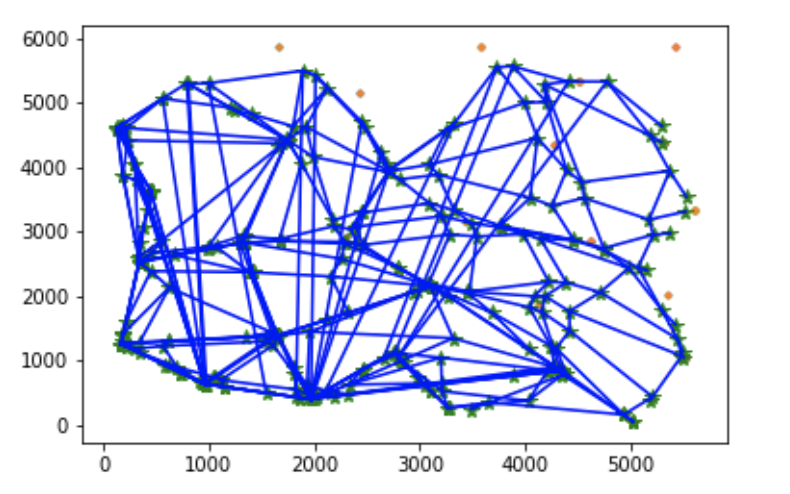
До кластеризации



После кластеризации

Слева - кластеризация K-means, справа - кластеризация SOM



SOM карта

**Сравнительная таблица 2.**

|  |  |  |
| --- | --- | --- |
| Метрика | K-means | Метод SOM |
| Rand Index | 1.0 | 1.0 |
| F1-measure | 1.0 | 1.0 |
| Fowlkes-Mallows Index | 1.0 | 1.0 |

**Пример 3.**

4 линейно разделимых класса, находящихся близко друг к другy

Параметры:

k=4

som\_size=20

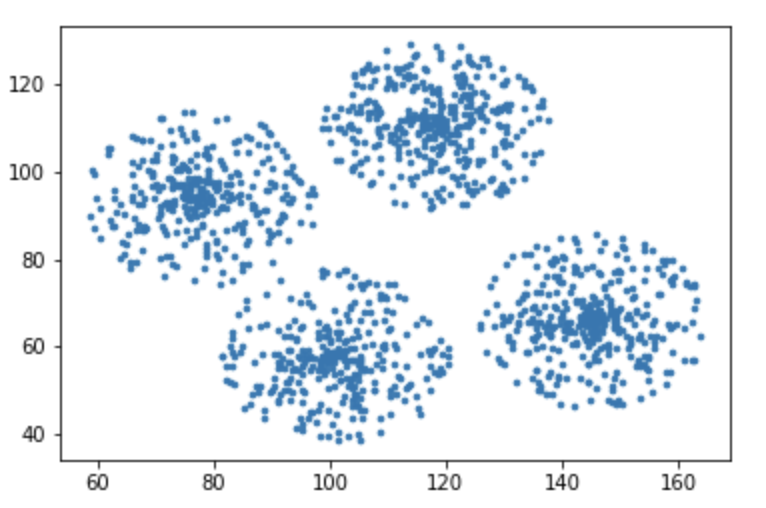
dimension = 2

learning\_rate=0.1

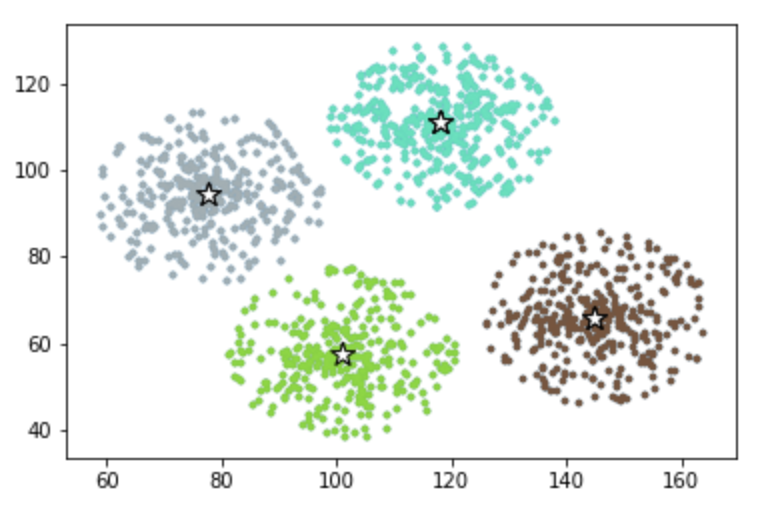
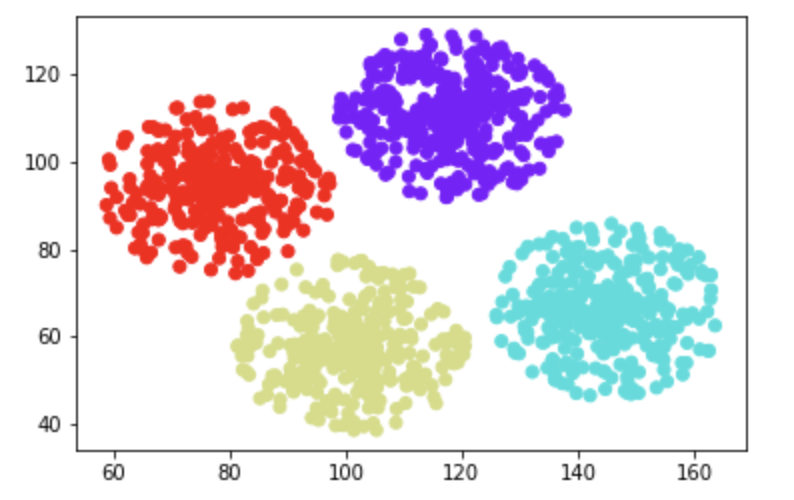
cooperation=8

epoch = 8000

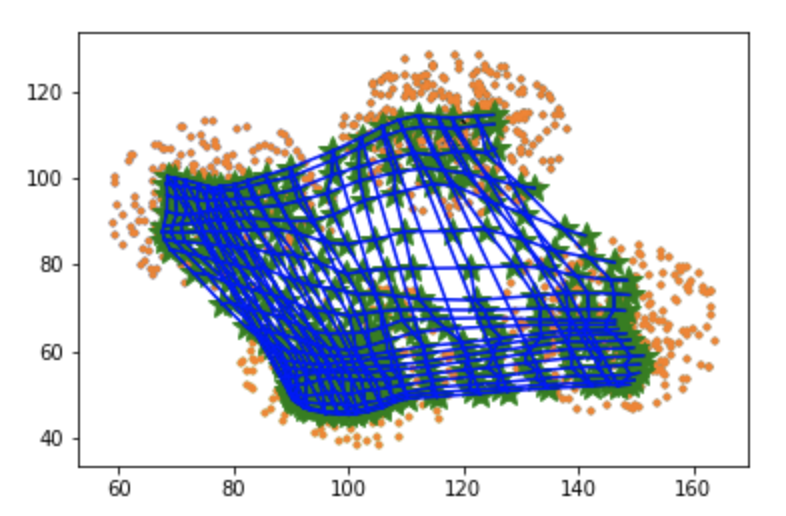
До кластеризации



После кластеризации

Слева - кластеризация K-means, справа - кластеризация SOM



SOM карта

**Сравнительная таблица 3.**

|  |  |  |
| --- | --- | --- |
| Метрика | K-means | Метод SOM |
| Rand Index | 0.831966893105333 | 1.0 |
| F1-measure | 0.7106256659551798 | 1.0 |
| Fowlkes-Mallows Index | 0.7172020757222693 | 1.0 |

**Пример 4.**

4 линейно неразделимых класса, средняя площадь пересечения классов 10-20%

Параметры:

k=4

som\_size=15

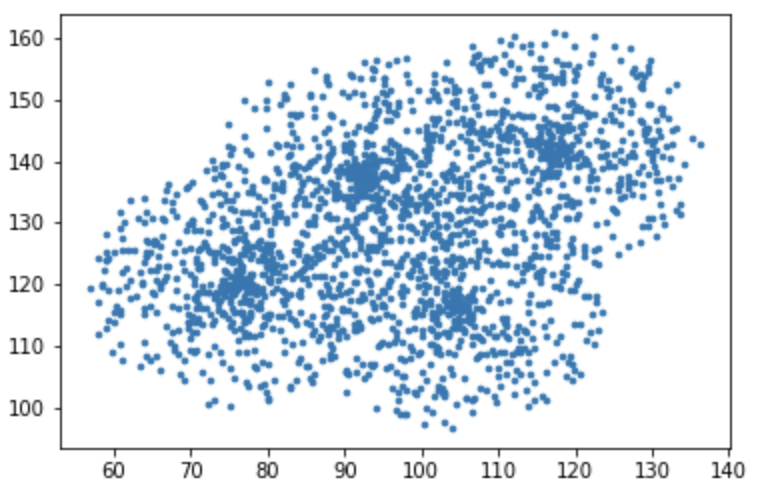
dimension = 2

learning\_rate=0.1

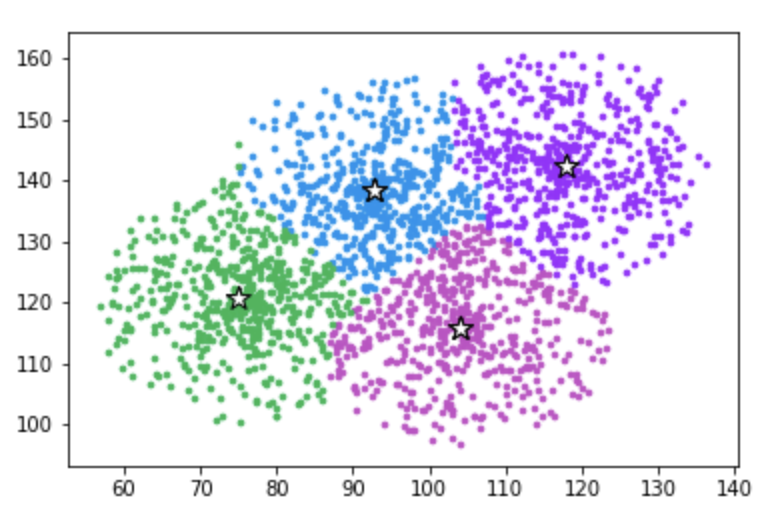
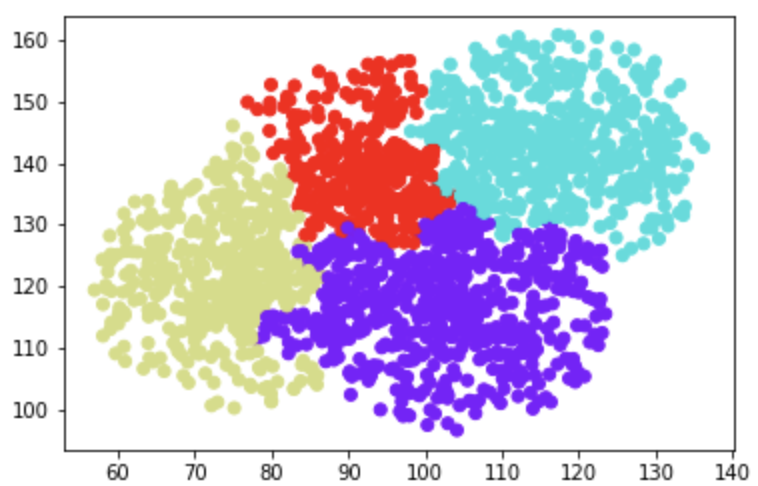
cooperation=5

epoch = 5000

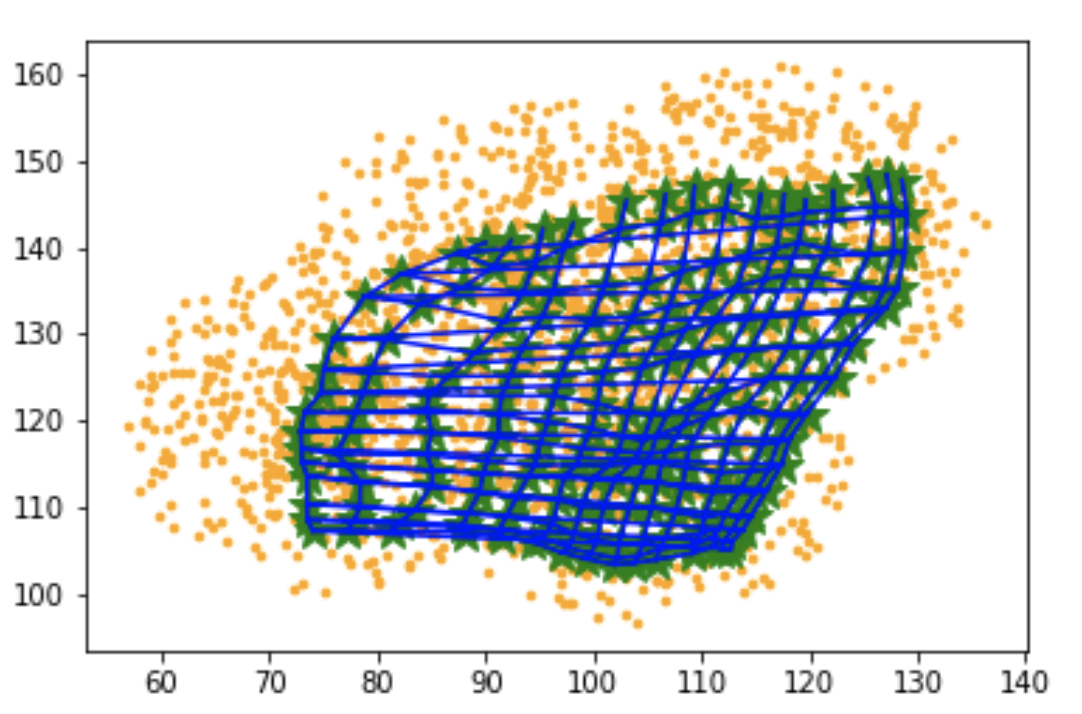
До кластеризации



После кластеризации

Слева - кластеризация K-means, справа - кластеризация SOM



SOM карта

**Сравнительная таблица 4.**

|  |  |  |
| --- | --- | --- |
| Метрика | K-means | Метод SOM |
| Rand Index | 0.8704550145228112 | 0.8474702118769916 |
| F1-measure | 0.740702489747801 | 0.6985004875654233 |
| Fowlkes-Mallows Index | 0.7407025017054031 | 0.6985531913370527 |

**Пример 5.**

10 линейно неразделимых класса, средняя площадь пересечения классов 50-70%

Параметры:

k=20

som\_size=25

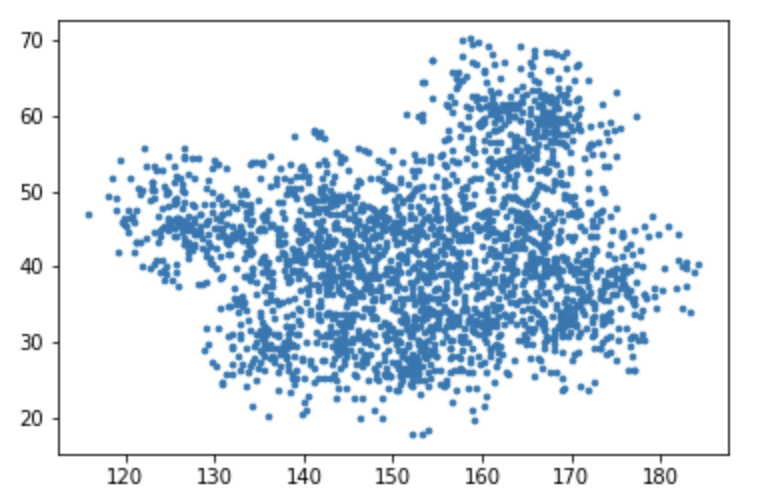
dimension = 2

learning\_rate=0.1

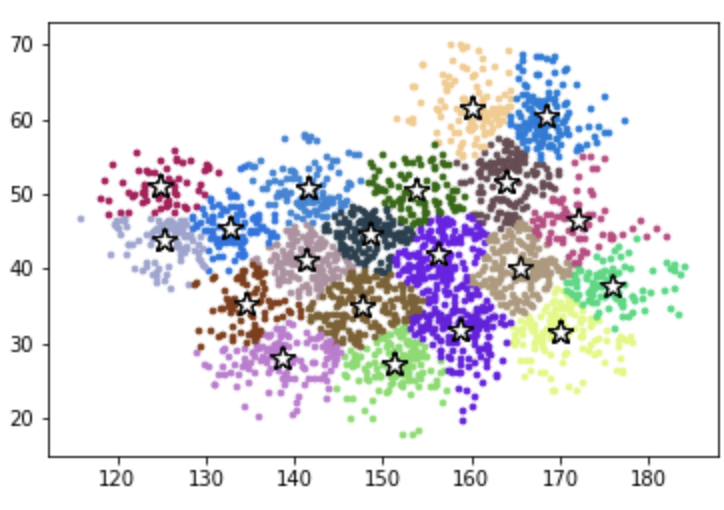
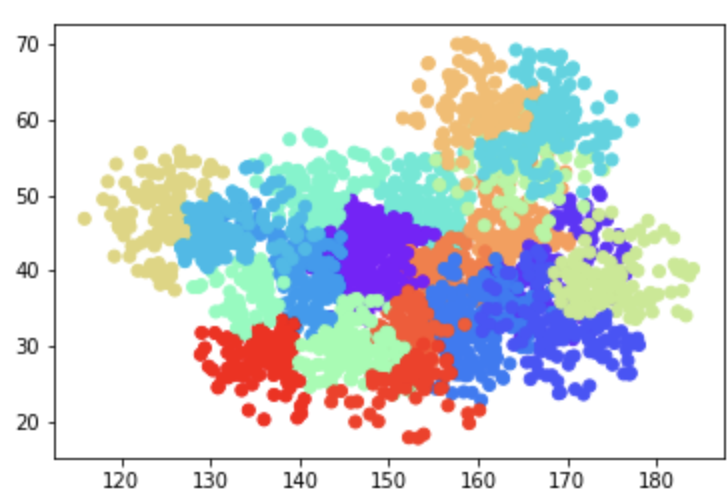
cooperation=8

epoch = 15000

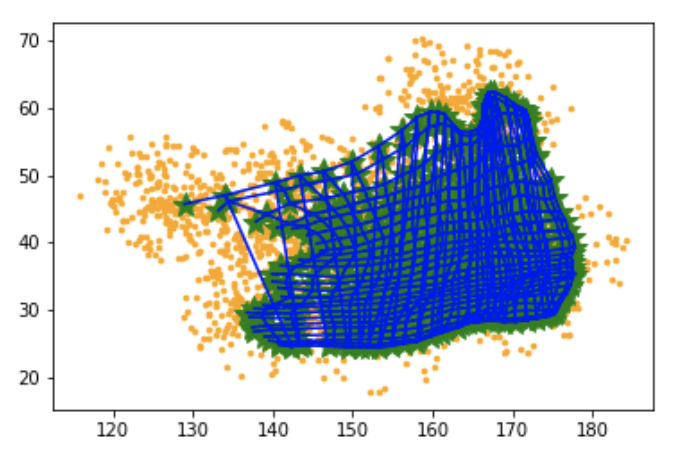
До кластеризации



После кластеризации

Слева - кластеризация K-means, справа - кластеризация SOM



SOM карта

**Сравнительная таблица 5.**

|  |  |  |
| --- | --- | --- |
| Метрика | K-means | Метод SOM |
| Rand Index | 0.9338668349192536 | 0.9499912098549216 |
| F1-measure | 0.35920836238573944 | 0.5211566528471501 |
| Fowlkes-Mallows Index | 0.3593038790920374 | 0.5214684174673907 |

**Пример 6.**

2 эталонных датасета с различным количеством признаков, но не менее 7

K-means – евклидова метрика

Wine Data Set – 3 класса, 13 признаков

Rand Index: 85.23%

Fowlkes-Mallows Index: 90.17%

Homogentity score: 84.53%

**Пример 7.**

Breast cancer dataset – 2 класса, 31 признак

Rand Index: 99.29%

Fowlkes-Mallows score: 99.67%

Homogentity score: 98.16%