

Министерство образования и науки Российской Федерации
Федеральное агентство по образованию
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Вятский государственный университет»

Факультет автоматики и вычислительной техники

Кафедра электронных вычислительных машин

Отчет по лабораторной работе №3-4 дисциплины
«Системы обработки знаний»

Выполнил студент группы ИВТ-41 _____/Крючков И. С./
Проверил _____/Ростовцев В. С./

1. Цель

Ознакомиться с основными командами создания, обучения и применения нейронных сетей кластеризации и классификации данных в Neural Network Toolbox с помощью сетей Кохонена и LVQ

2. Задание

- 1) Выполнить кластеризации с применением нейронных сетей Кохонена в соответствии с заданием (Приложение А). Создать и выполнить моделирование нейронной Кохонена, согласно согласованному с преподавателем, приведенному в приложении А. Результаты кластеризации вывести в виде графика с разными цветами кластеров, а также точки тестовых векторов.
- 2) Для создания сети LVQ обучающую последовательность использовать сеть Кохонена.

3. Ход работы

3.1. Самоорганизующаяся карта Кохонена

3.1.1. Обучение

Работа будет вестись с набором данных, в котором хранится информация о длине клюва и массе пингвинов. Датасет приведен в приложении А.

Отобразим элементы выборки с помощью следующего кода:

```
plot(penguins(:, 1), penguins(:,2), '.r', 'MarkerSize', 20)
hold on
```

Результат выполнения команд представлен на рисунке 1.

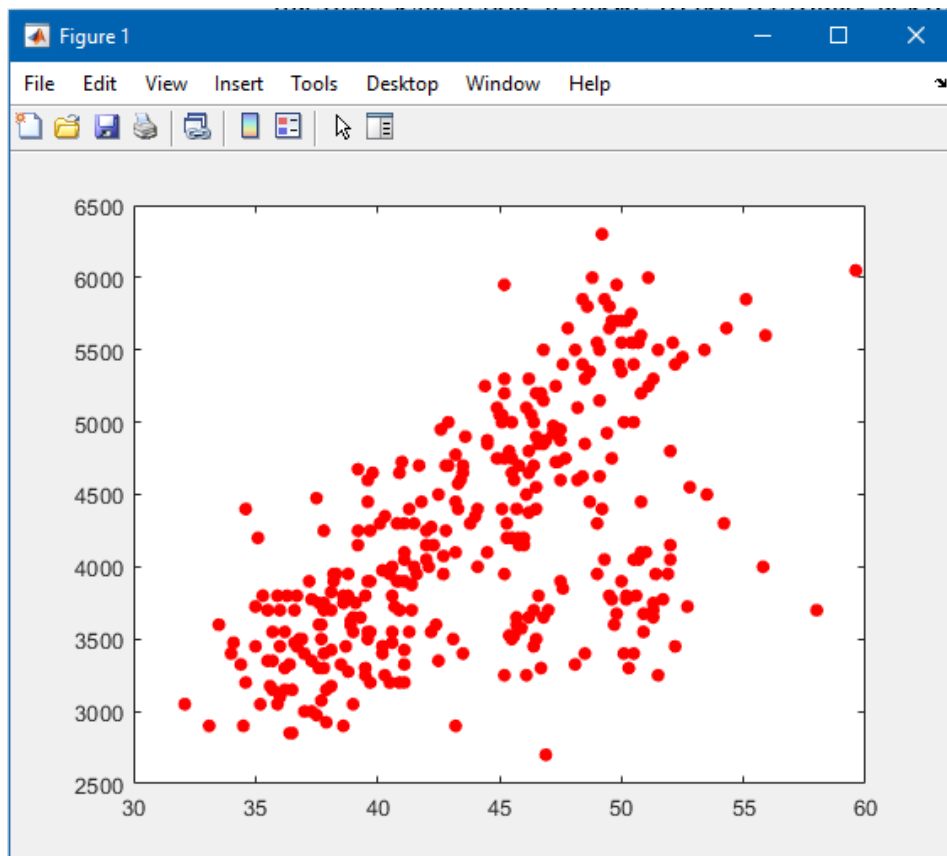


Рисунок 1 – Отображение элементов обучающей выборки

Создание и обучение самоорганизующейся карты Кохонена выполняется следующими командами:

```
net = selforgmap([6,6]);  
net = train(net,penguins');
```

Результат обучения представлен на рисунке 2.

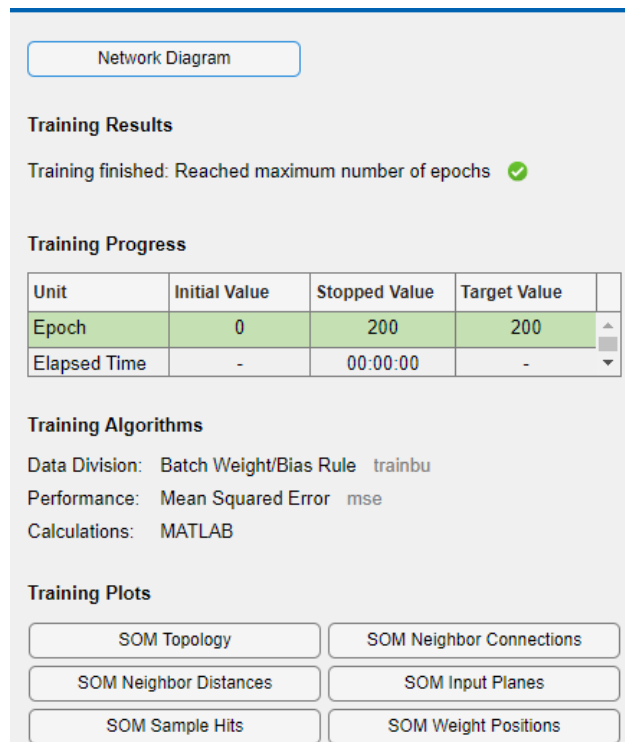


Рисунок 2 – Результат обучения сети

Структура сети представлена на рисунке 3.

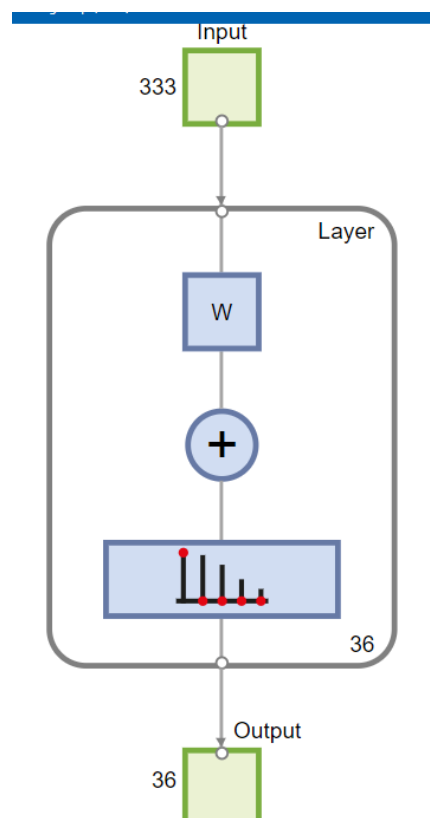


Рисунок 3 – Структура сети

Узлы карты Кохонена отображены на графике с помощью кода:

```
W = net.iw{1,1}
length(W)
```

```

for i = 1:length(W)
    text(W(i,1),W(i,2),int2str(i),'HorizontalAlignment','center',
        'VerticalAlignment','bottom')
end
hold on

plotsom(net.iw{1,1},net.layers{1}.distances)
set(gca, 'DataAspectRatio', [1, 200, 1])

hold off

```

Результат выполнения кода представлен на рисунке 4.

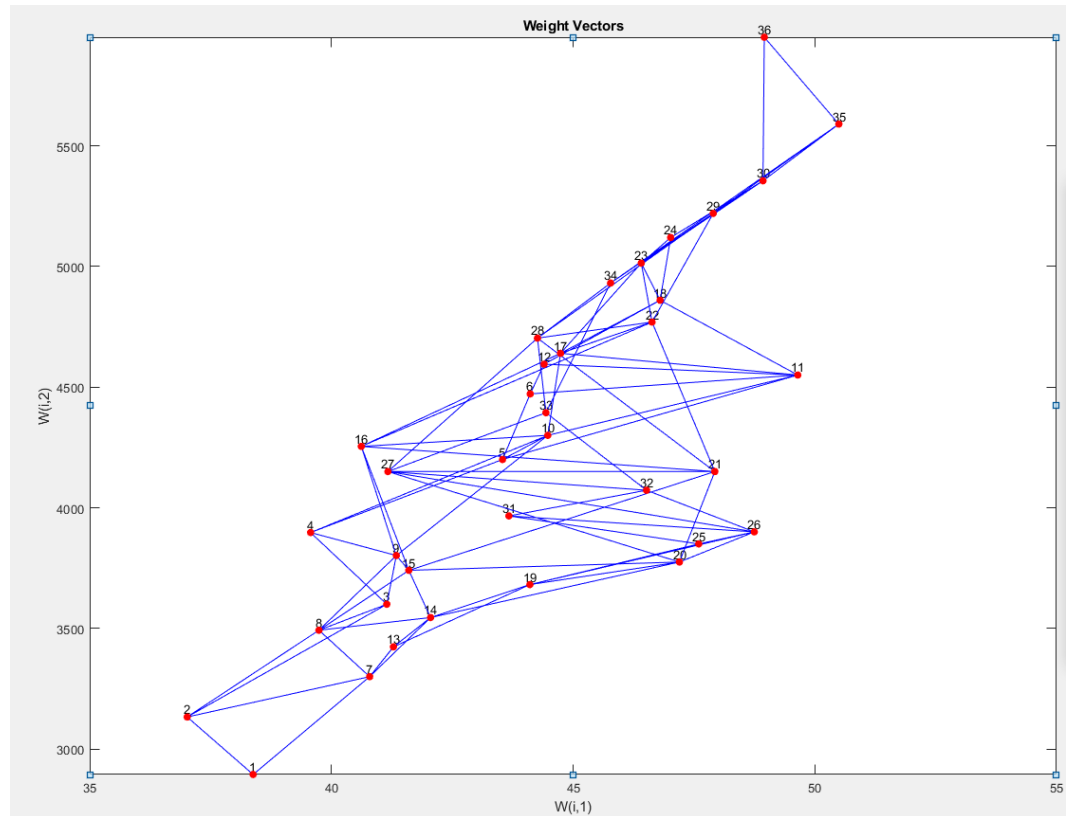


Рисунок 4 – Узлы карты Кохонена

Как видно из полученного графика, были выделены кластеры, размером с заданной картой.

Распределение элементов выборки по кластерам приведено на рисунке 5.

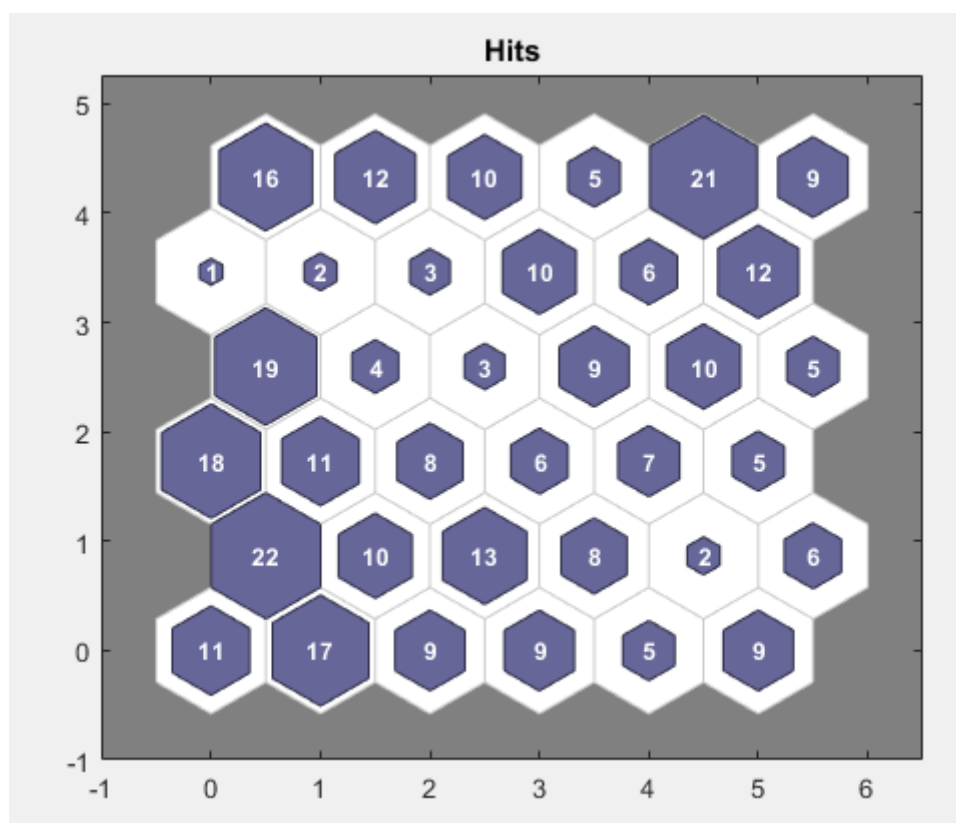


Рисунок 5 – Распределение элементов выборки по кластерам

3.1.2. Тестирование сети Кохонена

Для тестирования необходимо выбрать несколько примеров, которые не участвовали в обучении. Тестовая выборка представлена в таблице 1.

Таблица 1 – Тестовая выборка

Длина клюва, мм	Масса, г
50,4	5750
45,2	5200
49,9	5400

Отобразим тестовые примеры на графике зеленым цветом – рисунок 6.

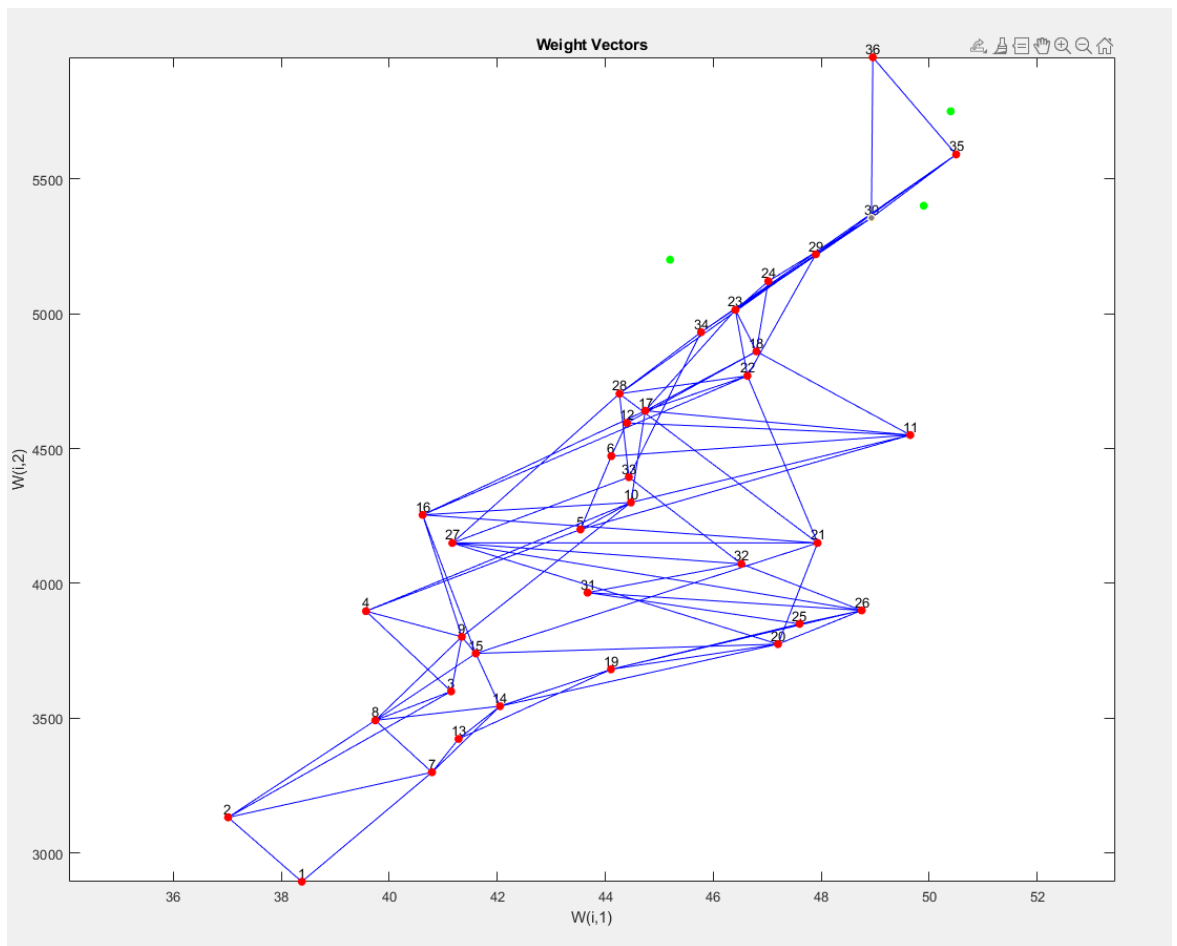


Рисунок 6 – Расположение тестовых примеров

Для тестирования выполним следующий код:

```
plotsomhits(net, TestData')
```

Результат выполнения представлен на рисунке 7.

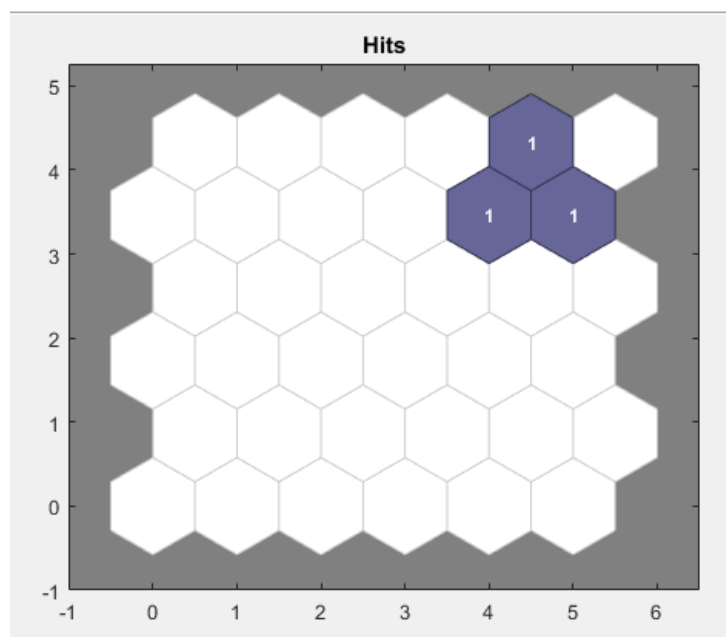


Рисунок 7 – Распределение элементов тестовой выборки по кластерам

По рисунку 7 видно, что в ходе тестирования элементы выборки были отнесены к соседним кластерам. Соседние элементы обычно классифицируют похожие образцы.

3.1.3. Выводы по сети Кохонена

В данном разделе была создана сеть Кохонена с помощью программы Matlab. Обучение сети выполнялось с помощью выборки, содержащей данные о длине клюва и массе пингвинов. В результате обучения сети, элементы выборки были разбиты на определенные кластеры, что показано на графиках. Так же было выполнено тестирование сети, в результате которого элементы выборки были отнесены к кластерам, наиболее соответствующим их значениям.

3.2. Сеть LQV

3.2.1. Обучение

Обучение сети LVQ будет выполняться с использованием результатов сети Кохонена.

Разобьем входную последовательность на 6 классов, результат представлен на рисунке 8.

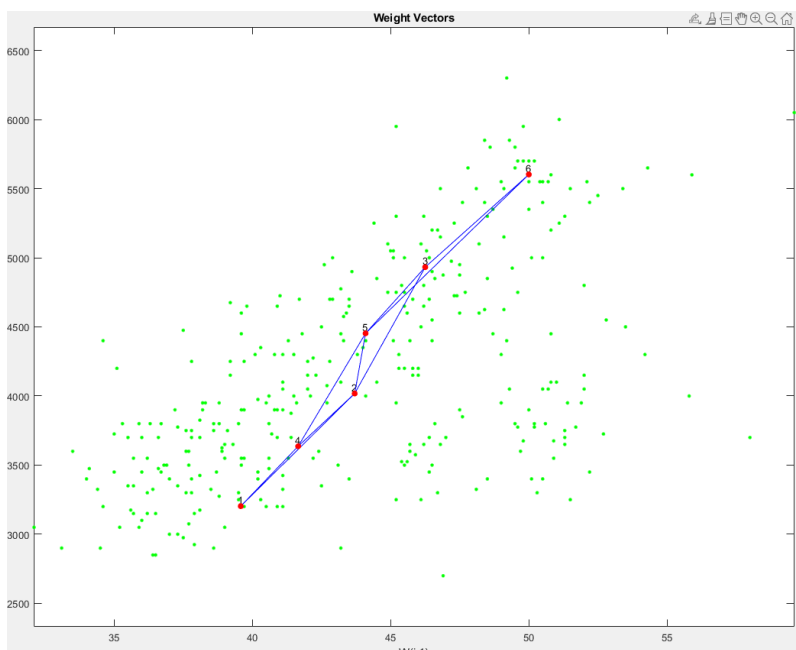


Рисунок 8 – Разбиение на классы

Заменяем точки числами их классов, также зададим им разный цвет следующим кодом:

```
tpenguins = penguins';
T = [];
for i = 1:length(tpenguins)
    y = net([tpenguins(1, i); tpenguins(2, i)]);
    cluster_index = vec2ind(y);
    T = [T cluster_index];
end
colors = ['r', 'b', 'g', 'c', 'm', 'y'];
figure(1), clf, axis([min(tpenguins(1,:))-5,max(tpenguins(1,:))+5,min(tpenguins(2,:))-5,max(tpenguins(2,:))+5]), hold on
for i = 1:6
    tmp = find(T==i);
    s = num2str(i) + 'r';
    text(tpenguins(1,tmp),tpenguins(2,tmp), num2str(i),'Color',colors(i));
    hold on
end
```

Результат представлен на рисунке 9.

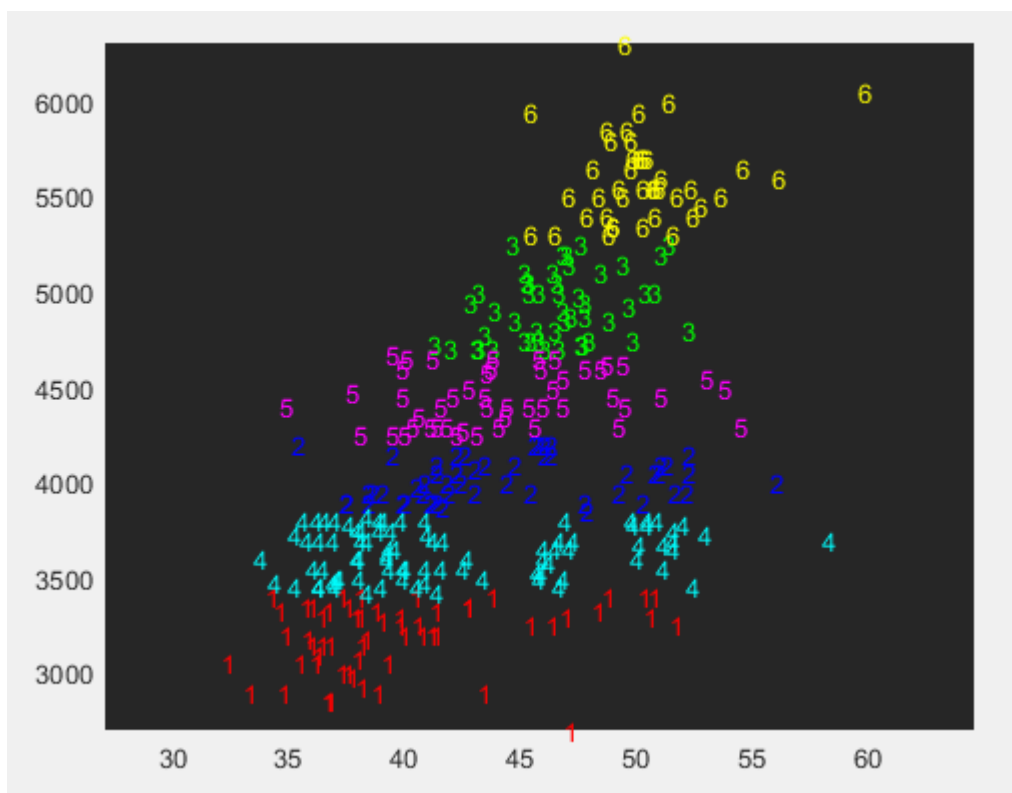


Рисунок 9 – Визуальное представление классов

Код для создания и обучения сети LVQ приведен ниже:

```
t = ind2vec(T);
lnet = lvqnet(8,0.2,'learnlv2');
lnet.trainParam.epochs=200;
lnet=train(lnet,tpenguins,t);
```

Результат обучения сети представлен на рисунке 10.

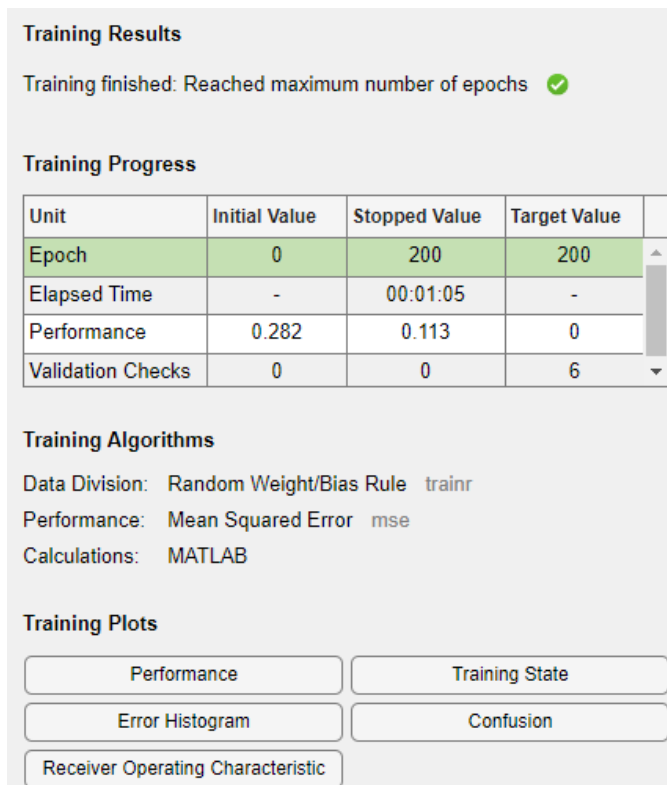


Рисунок 10 – Результат обучения сети

Структура сети представлена на рисунке 11.

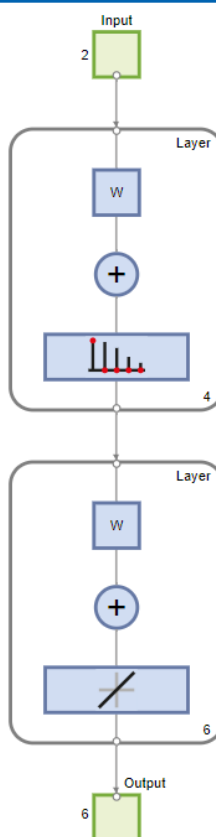


Рисунок 11 – Структура сети

После обучения получаем следующие весовые коэффициенты нейронов конкурирующего слоя, которые определяют положения центров кластеризации:

```
1.0e+03 *  
  
0.0437    3.2737  
0.0531    3.9455  
0.0424    4.8263  
0.6330    4.4343  
0.5564    4.4595  
0.0103    4.6141  
0.5999    4.4663  
0.0509    5.5548
```

Отообразим их на графике (рисунок 11, цвет точек центров не зависит от цветов изначальных кластеров):

```
plotvec(lnet.IW{1}',vec2ind(lnet.LW{2}),'o');
```

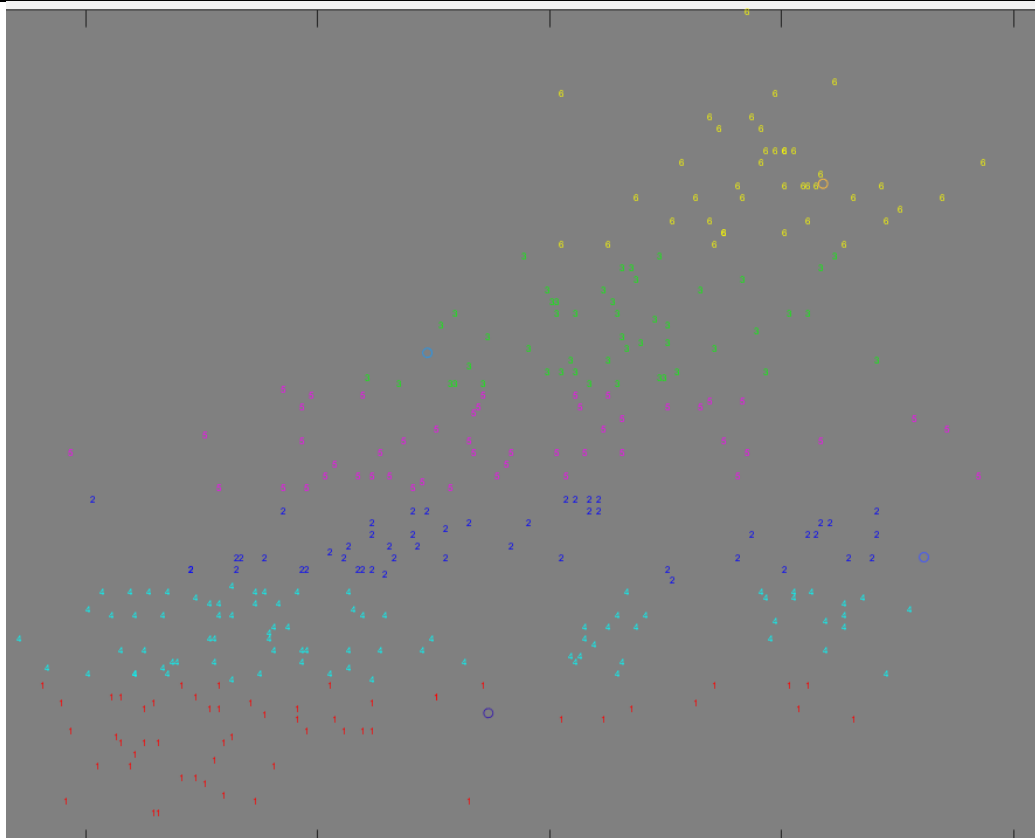


Рисунок 11 – Центры кластеризации

3.2.2. Тестирование

Тестовые примеры приведены в таблице 2.

Таблица 2 – Тестовая выборка для LVQ

Длина клюва, мм	Масса, г
50,4	5750
45,2	5200
51,5	5500

Тестирование показано на рисунке 12

```
>> c1 = vec2ind(1net([50.4, 5750]'))

c1 =

     6

>> c1 = vec2ind(1net([45.2, 5200]'))

c1 =

     3

>> c1 = vec2ind(1net([51.5, 5500]'))

c1 =

     6
```

Результаты показали, что кластеры тестовой выборки соответствуют изначальному распределению, показанному на графике 9.

3.2.3. Выводы по сети LVQ

В данном разделе была обучена сеть LVQ в программе Matlab. Сеть LVQ требует указания классов во входной выборке, поэтому для обучения сети LVQ были взяты результаты обучения сети Кохонена. Было выполнено тестирование на данных, которые не участвовали в обучении сети.

Выводы

В ходе выполнения лабораторной работы была решена задача кластеризации путем создания и моделирования нейронных сетей Кохонена и LVQ на заданном наборе исходных данных в программе Matlab. Результаты кластеризации и классификации зависят от количества кластеров, размерности входных данных и их качества, количества эпох, указанное при обучении сети Кохонена.

Приложение А.

Датасет.

Источник: <https://www.kaggle.com/datasets/youssefaboelwafa/clustering-penguins-species/data>

Длина клюва, мм	Масса, г
39.1	3750
39.5	3800
40.3	3250
36.7	3450
39.3	3650
38.9	3625
39.2	4675
34.1	3475
42	4250
37.8	3300
37.8	3700
41.1	3200
38.6	3800
34.6	4400
36.6	3700
38.7	3450
42.5	4500
34.4	3325
46	4200
37.8	3400
37.7	3600
35.9	3800
38.2	3950
38.8	3800
35.3	3800
40.6	3550
40.5	3200
37.9	3150
40.5	3950
39.5	3250
37.2	3900
39.5	3300
40.9	3900
36.4	3325
39.2	4150
38.8	3950
42.2	3550
37.6	3300
39.8	4650

36.5	3150
40.8	3900
36	3100
44.1	4400
37	3000
39.6	4600
41.1	3425
37.5	2975
36	3450
42.3	4150
39.6	3500
40.1	4300
35	3450
42	4050
34.5	2900
41.4	3700
39	3550
40.6	3800
36.5	2850
37.6	3750
35.7	3150
41.3	4400
37.6	3600
41.1	4050
36.4	2850
41.6	3950
35.5	3350
41.1	4100
35.9	3050
41.8	4450
33.5	3600
39.7	3900
39.6	3550
45.8	4150
35.5	3700
42.8	4250
40.9	3700
37.2	3900
36.2	3550
42.1	4000
34.6	3200
42.9	4700
36.7	3800
35.1	4200
37.3	3350
41.3	3550

36.3	3800
36.9	3500
38.3	3950
38.9	3600
35.7	3550
41.1	4300
34	3400
39.6	4450
36.2	3300
40.8	4300
38.1	3700
40.3	4350
33.1	2900
43.2	4100
35	3725
41	4725
37.7	3075
37.8	4250
37.9	2925
39.7	3550
38.6	3750
38.2	3900
38.1	3175
43.2	4775
38.1	3825
45.6	4600
39.7	3200
42.2	4275
39.6	3900
42.7	4075
38.6	2900
37.3	3775
35.7	3350
41.1	3325
36.2	3150
37.7	3500
40.2	3450
41.4	3875
35.2	3050
40.6	4000
38.8	3275
41.5	4300
39	3050
44.1	4000
38.5	3325
43.1	3500

36.8	3500
37.5	4475
38.1	3425
41.1	3900
35.6	3175
40.2	3975
37	3400
39.7	4250
40.2	3400
40.6	3475
32.1	3050
40.7	3725
37.3	3000
39	3650
39.2	4250
36.6	3475
36	3450
37.8	3750
36	3700
41.5	4000
46.5	3500
50	3900
51.3	3650
45.4	3525
52.7	3725
45.2	3950
46.1	3250
51.3	3750
46	4150
51.3	3700
46.6	3800
51.7	3775
47	3700
52	4050
45.9	3575
50.5	4050
50.3	3300
58	3700
46.4	3450
49.2	4400
42.4	3600
48.5	3400
43.2	2900
50.6	3800
46.7	3300
52	4150

50.5	3400
49.5	3800
46.4	3700
52.8	4550
40.9	3200
54.2	4300
42.5	3350
51	4100
49.7	3600
47.5	3900
47.6	3850
52	4800
46.9	2700
53.5	4500
49	3950
46.2	3650
50.9	3550
45.5	3500
50.9	3675
50.8	4450
50.1	3400
49	4300
51.5	3250
49.8	3675
48.1	3325
51.4	3950
45.7	3600
50.7	4050
42.5	3350
52.2	3450
45.2	3250
49.3	4050
50.2	3800
45.6	3525
51.9	3950
46.8	3650
45.7	3650
55.8	4000
43.5	3400
49.6	3775
50.8	4100
50.2	3775
46.1	4500
50	5700
48.7	4450
50	5700

47.6	5400
46.5	4550
45.4	4800
46.7	5200
43.3	4400
46.8	5150
40.9	4650
49	5550
45.5	4650
48.4	5850
45.8	4200
49.3	5850
42	4150
49.2	6300
46.2	4800
48.7	5350
50.2	5700
45.1	5000
46.5	4400
46.3	5050
42.9	5000
46.1	5100
44.5	4100
47.8	5650
48.2	4600
50	5550
47.3	5250
42.8	4700
45.1	5050
59.6	6050
49.1	5150
48.4	5400
42.6	4950
44.4	5250
44	4350
48.7	5350
42.7	3950
49.6	5700
45.3	4300
49.6	4750
50.5	5550
43.6	4900
45.5	4200
50.5	5400
44.9	5100
45.2	5300

46.6	4850
48.5	5300
45.1	4400
50.1	5000
46.5	4900
45	5050
43.8	4300
45.5	5000
43.2	4450
50.4	5550
45.3	4200
46.2	5300
45.7	4400
54.3	5650
45.8	4700
49.8	5700
46.2	4650
49.5	5800
43.5	4700
50.7	5550
47.7	4750
46.4	5000
48.2	5100
46.5	5200
46.4	4700
48.6	5800
47.5	4600
51.1	6000
45.2	4750
45.2	5950
49.1	4625
52.5	5450
47.4	4725
50	5350
44.9	4750
50.8	5600
43.4	4600
51.3	5300
47.5	4875
52.1	5550
47.5	4950
52.2	5400
45.5	4750
49.5	5650
44.5	4850
50.8	5200

49.4	4925
46.9	4875
48.4	4625
51.1	5250
48.5	4850
55.9	5600
47.2	4975
49.1	5500
47.3	4725
46.8	5500
41.7	4700
53.4	5500
43.3	4575
48.1	5500
50.5	5000
49.8	5950
43.5	4650
51.5	5500
46.2	4375
55.1	5850
44.5	4875
48.8	6000
47.2	4925
46.8	4850
50.4	5750
45.2	5200
49.9	5400